

RESULT 6  
 TVCHLV  
 epidermal growth factor receptor precursor - chicken  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erdb  
 C:Species: Gallus gallus (chicken)  
 C>Date: 28-Feb-1986 #sequence\_revision 05-May-1995 #text\_change 04-Feb-2000  
 C/Accession: A27720; A00643  
 R/Lax, I.; Johnson, A.; Howk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennart  
 Mol. Cell. Biol. 8, 1970-1978, 1988  
 A/Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mol  
 A/Reference number: A27720; MUID:88261272; PMID:3260329  
 A/Accession: A27720  
 A/Molecule type: mRNA  
 A/Residues: 1-1223 <LAX>  
 A/Cross-references: GB:M20366  
 R/Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M  
 Cell 41, 719-726, 1985  
 A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pcr  
 A/Reference number: A00643; MUID:85228222; PMID:2988784  
 A/Accession: A00643  
 A/Molecule type: mRNA  
 A/Residues: 585-1223 <NIL>  
 A/Cross-references: GB:M10066  
 C/Genetics:  
 A/Gene: erbB  
 C/Superfamily: epidermal growth factor receptor: protein kinase homology  
 C/Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor  
 C/peptide: protein kinase  
 F:1-30/Domain: signal sequence #status predicted <SIG>  
 F:31-1223/Product: epidermal growth factor receptor #status predicted <MAT>  
 F:31-654/Domain: extracellular #status predicted <EXT>  
 F:81-307/Domain: EGF receptor extracellular domain repeat <EE1>  
 F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>  
 F:655-677/Domain: transmembrane #status predicted <TM>  
 F:678-1223/Domain: intracellular #status predicted <INT>  
 F:719-984/Domain: protein kinase homology <KIN>  
 F:727-735/Region: protein kinase ATP-binding motif  
 F:136-202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #  
 F:192,650/Binding site: carbohydrate (Ser) (covalent) #status predicted  
 F:667/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:754/Active site: Lys #status predicted  
 F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query March 44.6%; Score 3053.5; DB 1; Length 1223;  
 Best Local Similarity 48.2%; Pred. No. 1.7e-119;  
 Matches 631; Conservative 169; Mismatches 347; Indels 161; Gaps 29;

QY 8 RWGLLALLPPGAA-----STQVCTGTMKLRLPASPETHLMDLRHLVGGCQVYQGNLE 61  
 DB 13 RGAAVLVLLLLGVALGSAVEEKVCOCGNNKLTOLGHVEDHFTSLQGMANNCEVLSNLE 72  
 QY 62 LTVPTNASISFLDIOEVOGYVLIANNOVROVLOLRIVRGQLFEDNYALAVLNDGD 121  
 DB 73 ITVEHNRDLFTLTIOEVAGYVLIANMDVLEMLQIRGNVLTDSFALAVSNVH 132  
 QY 122 PLNNTTVTGASPGELRELQSLTEILKGVLIQNPOLCYDPTILMKDIFKNNQAL 181  
 DB 133 -MNTQ-----GLRELPMKRLSEILNGVKISNNPKLCNMDTVLMDNIIDTSRK-PL 182  
 QY 182 TLID-TNRSAACHPCSPMGKSGRWGSSSEDCSLTTCVAGGA-RCKRPLTPDCHEQ 239  
 DB 183 TVLDFASNLSCPCPHNCTEDHGMAGBQNCQTLTVICAQOQSGCRKQVSDCHNQ 242  
 QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVYNTDTESPMPREGRYTFASCVTAC 299  
 DB 243 CAAGCTGPRSDCLACKKFRDADCTCPTPLVYNTTYQMDNPNRGKTSFATCTREC 302  
 QY 300 PYNLTSDVGSCTIVCLAHNOEVAHGTORCEKSGPCARVCTGLGQYIKANSKFIGI 359  
 DB 303 PHNVVTVDHGSCVASCNTDITYEV-EENGVRKCKCKDGLGKVCNIGTIGELKGLIS-INA 360  
 QY 360 TELE-FAGCKKIFGSLAFLESFDDPASNTAPLQBPQLQVFETLEITGYLVISAMPDS 418

DB 361 TNIDSPKCTKINGDVSLPVAFLDGAFKTLPLDPKLDVFRVVEISGFLLIQMPON 420  
 QY 419 LPLDSVFNQLVIRGRILHNCAVSLTLQGLISWLGRLRELGSGLALIHNTHLCPVH 478  
 DB 421 ATDLVAFENLEIRGTTRKHQGYSLAVNVLKQSLGRLKLSDDIMMKNKLCYAD 480  
 QY 479 TVPMDLFPNPHQALHTANRPDECEGGLACHQICARGHCWGPPTQCVNCSQFLRQ 538  
 DB 481 TMMRSLFATQSQKTKIIONRNNDCTADRHVCDPLCSGVGCMGPGPFHFCSCRFPSRK 540  
 QY 539 ECHEBRVIOGLPREVYNAHCLPCPREQPNQ---STCGPRADQVCAAHYDPPF 595  
 DB 541 ECVQGNILQGEFERERDSKCLPCHEBCLVNSTAVNTTCSGPFHDMCKAHFTDGFH 600  
 QY 596 CVARCPF---NNFVSWLRVPKVSASHLEBDEGACPCPINTGSCVDDDKCPAE 651  
 DB 601 CVACPGAVLGENDTL-VWKYA-----DANAVCQLHPNCTRGCGGLECP-- 647  
 QY 652 QRASPLTSIVAVY-GILLVVLGVVFGILIRROOKIKYTMRLLOTELVEPLTPSG 710  
 DB 648 -NGSKTPSIAGVVGGLCLVVGGLIGLYLRH-HVAKTLRLLORELEVEPLTPSG 705  
 QY 711 AMPNQAKRLKETELRKVYLGSGAFVYKGMIPDGENYKIPVAIKVLRNTSPKRN 770  
 DB 706 EAPNOAHLRIKETEFKTKVYLGSGAFVYKGMIPDGENYKIPVAIKVLRNTSPKRN 765  
 QY 771 KEILDAVYMAVGSPYVRLIGICTSTVQVLTQDLMYGLCLLDHVENRGRLGSDLLN 830  
 DB 766 KEILDAVYMAVGSPYVRLIGICTSTVQVLTQDLMYGLCLLDHVENRGRLGSDLLN 825  
 QY 831 WCMQIAKMSYLEDVLYNRDLAARVLYKSNHYKITPGLARLLDIDETVYHAGGV 890  
 DB 826 WCMQIAKMSYLEDVLYNRDLAARVLYKSNHYKITPGLARLLDIDETVYHAGGV 885  
 QY 891 PIKMALESILRRFTQSDVSYGVYVWELMTFGAKPYDGIPIREIPDLKEGERLP 950  
 DB 886 PIKMALESILRRFTQSDVSYGVYVWELMTFGAKPYDGIPIREIPDLKEGERLP 945  
 QY 951 PICTIVYIMYKCMWIDSECPRRRELVESEKMAKDPQRFVITQ-NEDLGPASPLDST 1009  
 DB 946 PICTIVYIMYKCMWIDSECPRRRELVESEKMAKDPQRFVITQ-NEDLGPASPLDST 1005  
 QY 1010 FRSILLEDMDGLVDAEFLVLPQGFPCPDPAAGVHHRHSSSTRSGGDTLTL 1069  
 DB 1006 FRSILLEDMDGLVDAEFLVLPQGFPCPDPAAGVHHRHSSSTRSGGDTLTL 1038  
 QY 1070 EPSEEBAPRSP-----APSEAGASDVFDGDLGMAKGLQSLPTHDPSPLOKYSDDPTV 1124  
 DB 1039 -----SRTPLSSLSATSNNSATNCID-----RNGQCHPVREDFVQYSSDPTG 1083  
 QY 1125 PLPSER-DGYVAPLTCSPQPEYVNOQDVRPQPSPRESPRLPAARPAAGTLERAKTSLFG 1182  
 DB 1084 NFLEESIDDFL-----PAPRYVNO-LMPKPS-----TANV 1114  
 QY 1183 KNGVVDVF-----AFGAVENPEVLTPOGAAPOPHPAPSPAFDNLV 1227  
 DB 1115 QNGIVNINSLTAISKLPMSORVQNSHSTAVNDPEL-----NTNGSLAKTYFESSP 1166  
 QY 1228 YMDQ-----DPE-----KQAPSTFGTPTAENPEVLTGLDVP 1260  
 DB 1167 YMIQGNQINLNDNDYODFLPNETKPNGLKVAENPEVLTGLDVP 1214

RESULT 7  
 A47253  
 epidermal growth factor receptor, HER4 - human  
 C/Species: Homo sapiens (man)  
 C/Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999  
 C/Accession: A47253  
 R/Plowman, G.D.; Culionescu, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L  
 Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993  
 A/Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide

A:Reference number: A47253, PMID:93189574; PMID:8383326  
A:Accession: A47253  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1308 <P>  
A:Cross-references: GB:107868; NID:q337359; PIDN:AA859446.1; PID:q337360  
A:Note: sequence extracted from NCBI backbone (NCBI:128842)  
C:Superfamily: epidermal growth factor receptor, protein kinase homology  
C:Keywords: ATP; growth factor receptor  
F:716-981/Domain: protein kinase homology <KIN>  
F:724-732/Region: protein kinase ATP-binding motif

Matches 568; Conservative 165; Mismatches 392; Indels 152; Gaps 31;

```

Qy 4 AALCRWGLLALLPPGAASL-----QVCTGDMKRLPLASPEHLDMLRHLYQGCQVQGN 59
Db 8 AALQO--LLVLISLRCCSTDPDRKVCQGTSMQMTM--LDNHLLKMKKKYSGCNVYLEN 62
Qy 60 LELTYLPTNASLSFLQDIQEVQGYVLLAHNOVROVPLQRLIVRGTOLEFEDNALAVLDN 119
Db 63 LEITYTOENQDLSPFQSIQEVQGYVLLAHNEVSTIPLVNLRLIRGQMLYBENFTLLWMSN 122
Qy 120 GDPINNTPTVPGASPGRLREQLRLSLEILKGVLIQIRNPOLQVODITLMKIDPHKNNQL 179
Db 123 YQK-NPSP--DYVQGLKQQLQSLNTEILSGYKVSINPLLCVETIIMMDYDKSNP 179
Qy 180 ALTLIDNRSRACHPCSPMCKSGRQWGSSEDCSLRTVYAGGC-ARCKGPLPTDCHE 238
Db 180 TMLNLPFAFERQCKCHGCVNCSQWAPFGHCQKFTLLCAEQCNCRGCPKPIDCCNE 239
Qy 239 QCAAGCTGPKHSDCLACHFNHSGICELHCPALVTYMTDFESMPNREGRTFGASCUTA 298
Db 240 HGAAGCTGPRATDCLACDFDDGTCOTCPKTIYDVSQVVDNPNIKITFGAACVKE 299
Qy 299 CPYVNLSTDVGSCTLVCPHNOEYTAEDGTQCEKCKSPCARVCYGLGMQYIKANSKFIG 358
Db 300 CPNVVYTE-GACVRSQAGMLEVD-ENGKRSCKPCQGVCPKVDGIGISL-SNTIAVN 356
Qy 359 ITEL-EPAGCKKIGSLAFLESTFDGDPASNTAFLQEBOLQVETLEITGYLYTAMPD 417
Db 357 STNRSFSCNCKINGDILINRNSFEGDPHYKIGTMDPEHMLNLTVEITGYLYTAMPPE 416
Qy 418 SLPLSVFQNTQVIRGRILHNGAYS-LTLQGLGISWELSGRLSGLALIHNTLHCF 476
Db 417 NMTLSVFNQLEITRGRTRFRSGFVYVYRHLQMLGRLSKVSAQNVLKXTLQRY 476
Qy 477 VHTVPMDQLEFNPHQALHTANRPDEBCEVGEGLACHQICARHGWGFPPTQVNCQGLR 536
Db 477 ANTIMRRLFRSEQDSIEYDART-----ENQTCNNECSDEGEM-DGPTMCVSCILHVD 528
Qy 537 GQECVEGRVYLOGPREVYVNRHLPCHEQCPQNGSVTGCPEADQVCAHAKXDPFC 596
Db 529 GGRVYASCNLQGEPRQVQVQCHQCELVDTSTICGPBPANCSAHQDQBPQC 588
Qy 597 VARP--FNNFTVSVFLVBPVKVASHLEPDEBACQPCPINTCHSCVDDDKCPAECR 653
Db 589 IPRCPHGLDGDTLWXYA-----DKMGQCCQCHQNCQGGSGPGLSCRD-1 637
Qy 654 ASPLTYSAVVGLLVVGLVGVGILIKRQCKIRKXTMARLLQELVELYEPPLPSGAMP 713
Db 638 VSHSLAVGLVSGLLITVALLVLLRRRIK-RKRTIRCLQEKELVEPLTPSQAP 696
Qy 714 NQAMRLIKETELRKVKLGSAGFTYKGIWIPDGENVKIPVAIKYLRNTSPKAKEI 773
Db 697 NQALRLIKETEFKDRYLSGAGFTYKGLMNDGENIRIPVAIKYLRNTSPKAKEI 756
Qy 774 LDEAYVWAGVSPVYSRLGICLTSTVQVLTQMPYGLLDHVENRGLSGQDLNWCW 833
Db 757 LDEAYVWASVDHPVCRLLGICLTSAVOLVTLQMPYGLLDYVQROHERICQGLNWCW 816
Qy 834 QIAKMSYLEDVRLVHRDLARVNLVKSFNHKTIDPGLARLLIDSTEYHADGKVPK 893
Db 817 QIAKMSYLEERHLVHRDLARVNLKPNHVKITDGLSKLLADEKEXYQADGKVPK 876
Qy 894 WMALESILRRRFTHQSDVSYGVTVWELMTFGAKPYDGIAPREIPDLLEKGERLLPQPI 953
Db 877 WMALESILQWYTHQSDVSYGVTVWELMTFGSKPYDGIAPKEIASYLENGERLLPQPI 936
Qy 954 TIDVYMTVMKQWMTDSECRPRFRRELVSFESMARDPQFVYVQVQEDGSPSPDSTYRS 1013
Db 937 TIEVYMTILKQWMTDPSRRPRFRRELVSFESMARDPSSYVLTQ--NLPSLSRRLFSR 993
Qy 1014 LLEDDMDGLVDAEYVLPQGGFCPPAPGAGGVHHRHRSSTRSGGDLTLGLEPSE 1073
Db 994 LLSDD--DVVDADEYLLPYKRI-----NRQGS----- 1019

```

RESULT 9

```

Qy 1074 BEAPRSPLPSEGCAGSDVEDGLGMGAAGLQSLPHPHDSPLQRYSEDPYV-PLPSETDG 1132
Db 1020 -----PCIPPTCH-----PVRENSITLRNISDPTQNALFXDLDG 1054
Qy 1133 YVAPPLTCSQPEPEVNPDPDRPQ-----PSPRE-----GLP-AARPAQALFPAKTL 1179
Db 1055 H-----EYVNPGETSRSRSLDIYVNPVEDLTQMGVPSLSSQEAETNFSPEYL 1104
Qy 1180 SPKNGVVDVFAFGAVENPEYLTPOGGAAPHPHPAPFAFDLLYWDODDPPERGAP 1239
Db 1105 NTQGNL-----PLVSSSMDDPY-----QAG-----YQAAF-----LPQTGL 1139
Qy 1240 PSTFKGPTPAENPEYLG 1256
Db 1140 TGNGMFLPAENLEYLG 1156

```

kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human

C/Species: Homo sapiens (man)

C/Date: 04-Oct-1991 #sequence\_revision 13-Jan-1993 #text\_change 17-Nov-2000

C/Accession: A36223; 159164

R/Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A. Proc. Natl. Acad. Sci. U.S.A. 86: 9193-9197, 1989

A/Title: Isolation and characterization of ERBB3, a third member of the ERBB/epidermal growth factor receptor gene family.

A/Accession: A36223; MUID:90083234; PMID:2687875

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-1342 <KRA>

A/Cross-references: GB:M29366

R/Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todarik Proc. Natl. Acad. Sci. U.S.A. 87: 4905-4909, 1990

A/Title: Molecular cloning and expression of another epidermal growth factor receptor gene.

A/Accession: A36223; MUID:90311312; PMID:2164210

A/Status: preliminary, translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>

A/Cross-references: GB:M34309; NID:9183990; PIDN:AAA5979.1; PID:9306841

C/Genetics:

A/Gene: GDB:ERBB3; HER3

A/Cross-references: GDB:119880; OMIM:190151

A/Map position: 12q13-12q13

C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase C; Key words: ATP; phosphotransferase

F/707-972/Domain: protein kinase homology <Kin>

F/715-723/Region: protein kinase ATP-binding motif

Query Match 34.8%; Score 2378.5; DB 2; Length 1342;

Best Local Similarity 39.8%; Pred. No. 1.7e-91;

Matches 526; Conservative 193; Mismatches 460; Indels 141; Gaps 34;

```

Qy 10 GLLALLPPGA--STQVCTGDMKRLPLASEPETHLDMRLHYQGCQVQGNLELYLPT 67
Db 11 GLLFSLARSGVNSGAVCPGLTNGLSVGAENQVQLYKLYERREVVMGNLELYLTH 70
Qy 68 NMSLSTLQDIQEVQGYVLLAHNOVROVPLQRLIVRGTOLEFEDNALAVLDNGDPINNT 127
Db 71 NADLSLQWIRETVGYLVANNEFSLTPPNIRYVRGTVQVYDGKFAIFW-----LNVNT 125
Qy 128 PVTGASPGGRLREQLRLSLEILKGVLIQIRNPOLQVODITLMKIDPHKNNQLALTDIN 187
Db 126 ----NSSHLRQLRLQLLEILSSGYIRKNDKLCMDITIMRDIYRD--AELVND 178
Qy 188 RSRACHPCSPMCKSGRQWGSSEDCSLRTVYAGGC-ARCKGPLPTDCHEQCAAGCTG 246
Db 179 NGRSCPCHGVCKG-RCWGPGSBDQTLTKTICAPQCNHCGFNPENQCHDEACAGCGG 237
Qy 247 PKSDCLACHFNHSGICELHCPALVTYMTDFESMPNREGRTFGASCUTAAPYVYST 306

```

```

Db      238 PDITDCAFACRHNFGAGACVPCRPDPLVYNNKLTFOLEPNPHTKYQYGVACVASCPHNFV-V 296
Qy      307 DVGSCITVCPHLHNOEVTAEADGTORCEKSKPCARCYGIGMVIYANKSF--IGITLE- 363
Db      297 DDTSCVACRCPDPMEDV-KNGLJMCPEPCGGLCPKACEGTG-----SSRFQTVDSNDIG 350
Qy      364 FAGCKKIFGSLAFLESPDGDASNTAPLOPELOVFTLEETIYLYISAMPDLPDS 423
Db      351 FVNCSTKIIGNLDFLTGLNGDPMHKKIPALDPEKLVNFRVREITGYLNIQSPHMHNS 410
Qy      424 VFOQLQVIRGRILHNGAYS-LTLQGLISWLGRLRELGSGLAIHNTHLCFHTATPW 482
Db      411 VSNLTJTGGRSLVNRGSLIMKLNNTSLGFRSLKETSAGRIYSANROLCYHHSILNW 470
Qy      483 DQLFNPHQALLHTA-NRPEDECVGGLACHQLCARGHGWGPGPTQCVNCSQFLRGCEV 541
Db      471 TVLRLGPTREERLDIKNNRRRRCVAGKVCDFLSSGGCGWGGPGQCLSCNYSRGVCV 530
Qy      542 EECRVYLCIPREYVNAHRLCPHPECOPONGSVTCFGEADOCVACAHYKDPFCVACP 601
Db      531 THCNFLNGEPREFAHAEFCFCHPECQMEGTATNGSGSDTCAQCAHFRDGPCHVSCP 590
Qy      602 FNNFTVFWLRVPKYSASHLEPDEGACQPCPINCSTHSCVDLDDKGCFAEQRA----SPL 657
Db      591 HGVLAGK-----GPYKYPDVQNECRPHCHENTQCGKELODCLGLVLIGKTHL 642
Qy      658 TSISAVAGILLVVVGAVFGLILRRQOKTR-KYWRRLLOETELVPLTPSGAMPNOA 716
Db      643 TMTLVIAQ--LVVIFMMIGFTFLYMRGRILNKRAMPRLYLERGESIPDLPS-EKAKV 699
Qy      717 QWRILKETELRKVKVLGSGAFGTYYKGIMIPGENVKKIPVMIKVLRENTSPKANEILDE 776
Db      700 LARIFEITELRKVLKLGSGVFTYKGVWIPGESIKIPVCIKVEDSGKROSPQAVDH 759
Qy      777 AYVMAVGSPVYSRLIGICLTSTVQLVTLQMPYGLLDHVRNENRGLSODILNMCQIA 836
Db      760 MLAGSLDAHIVRLGLGCPGSSLTQVLYPLGLSLDHVRHOGALPOLLIMWGVQIA 819
Qy      837 KMSYLEDVRLVHRLAARNLVKSPNHYKIIDPELARLDDIEHYADGKPYIKMA 896
Db      820 KMYTLEHGWHRNLAARNVLKSPVOVADFPVADLPPDDQLLYSEAKPIKMA 879
Qy      897 LESILRRFTHOSDVMSYGVTVWELMTFGAKPYDIPAREIPDLLEKGERLPQPICTID 956
Db      880 LESIHFGKXTHQSDVMSYGVTVWELMTFGAEFYALRLAEPVDLLEKGERLAQPICTID 939
Qy      957 VYMIWKKCMIDSECRPRRELVSFESMARDPQRFVYIQNEDLCPA---SPLDSTFYS 1013
Db      940 YVMWVKCMMDENIRPTEKELANFTMARDPREYLVYIKES-GRGIAFGPEPHGLTNK 998
Qy      1014 LLEDDMDGLVDAEEYLVPOQGFCCPDPAFGAGVWHRHSSSTRSGGDLTGLAP-S 1072
Db      999 KLEVEYLEBELDLDDLEAED-----NLATTIGSALSIPVGLYN 1039
Qy      1073 EEEARSPPLABEGAGSDVFGDGLMGAKGLQSLPTHD-PSPLORYSEDPVLP- 1127
Db      1040 RPRGOSLSPSSGY-MPMNQNLGESCQESAVSSSBRCPVSLH-----PMRSGCL 1092
Qy      1128 --SETDGYVA-----PLTCSPOE---YNNQPRVVRQPPSPRGP- 1162
Db      1093 ASESEGHVTSSEALQEKVSMCRSRSRSPRRPGDSAYHSORISLITPTVPLSPGLE 1152
Qy      1163 -----LPAARPAATLERAKTLP-GRKGV-----KDVAFGAVENPEYLPQG 1207
Db      1153 EEDVAGVMPDTHLKGTPSSREGTSSVGLSSVLTEBEDD-----DEYEMNR 1204
Qy      1208 GAAPQHPRPAPSPAFNDLYYWD-----QDPRGAPSTSTKGPITANPEYL 1255
Db      1205 RNSP-PHPRPSSLEELGYEYWDVGSDSLASIGSTQSCPLHPVIMPTAGTTPDEDEYEM 1263

```

RESULT 10  
JC4387

```

epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C>Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: J04387
C:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant pro
Gene 165, 279-284, 1995
A:Reference number: J04387; MUID:6096535; PMID:8522190
A:Accession: J04387
A:Molecule type: mRNA
A:Residues: 1-1339 <HEL>
A:Cross-reference: GB:U29339; NID:9915389; PID:9915390
A:Experimental source: liver
A>Note: The authors translated the codon AAC for residue 369 as Thr and GTT for re
C:Comment: This protein is a functional heregulin receptor that transduces signals
C:Genetics:
A:Gene: ErbB3
A:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase I
C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prot
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <Kin>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1219,1257,1259,1273,1286,1325/Binding site: phosphate (T
Query Match 33.5%; Score 2293.5; DB 2; Length 1339;
Best Local Similarity 40.1%; Pred. No. 5,7e-88;
Matches 518; Conservative 168; Mismatches 440; Indels 165; Gaps 35;
Qy      3 LAALCRWGLLALLPQCA--STQVCTGDMKRLPASETHLDMIRLYGCGQVYQGN 59
Db      7 LQVLC---FLTSARSEWNSQAVCFGLTNGLSTGDADNQQTLTKYKCEVVMGN 62
Qy      60 LELTYPTNASLSFLOIOGVGYVLIHNGVAVQVPLRRIYRGQLFEDNVALVLDN 119
Db      63 LEIVTGHANDLSFLQIRVTVYLVANNEFVPLNLRVVRGQVVDGKFAIFVM-- 120
Qy      120 GDLNNTPTPYTASPGGLARELQRLSTELIKGVLTIQRNPOCYODTILMKDI FHKNNQL 179
Db      121 ---LNVT---NSSHALRQLKFTQLTEILSGVYIEKNDLCHMDTIDWRDILVRV- 170
Qy      180 ALTLIDNRBRACHPGSPMKSGKRCMGESSEDDQSLTRYVCAGG-ARCKGRLPTCCHE 238
Db      171 GAEIVVKNNGANCPCHVEVCKG-RVWGPPDDQILLTKITCAQCGKRCGPRPNQCCHD 229
Qy      239 QCAAGCTGPRGSDCLACLHFNHSGICELHCRPALVTYNTDTFESMPNREGRYTFGASCVTA 298
Db      230 ECAGGGCGPQDTCFACRRRNDGACVPRREPLVYNNKLTFOLEPRPHTKYQYGVACV 289
Qy      299 CFYNYLSTDVSGCTVCPHLHNOEVTAEADGTORCEKSKPCARCYGL--GMQYIKANSKF 356
Db      290 CFHNFV-VDTQFCVACRCPDPMEDV-KHGLJMCPEPCGGLCPKACEGTGSSRYQTVDSN 347
Qy      357 IGITLEFAFCKKIFGSLAFLESPDGDASNTAPLOPELOVFTLEETIGLYISAMP 416
Db      348 ID-----GFVNCSTKIIGNLDFLTGLNDVPMHKKIPALDPEKLVNFRVREITGYLNIQSNP 403
Qy      417 DSLPDISVFONLQVIRGRILHNGAYS-LTLQGLISWLGRLRELGSGLAIHNTHLC 475
Db      404 PHMHNFVSFSLNTLTIGRSLVNRGSLIMKLNNTSLGFRSLKETSAGRIYSANROLCYH 463
Qy      476 FVHTVPMDOCFRPHQALLHTA-NRPEDECVGGLACHQLCARGHGWGPGPTQCVNCSQF 534
Db      464 YHSLNMTKRLRSPSEERLDIKYDRPLGELAGKVCDFLSSGGCGWGPAPGQCLSCNYS 523
Qy      535 LRQCEVCECRVYLCIPREYVNAHRLCPHPECOPONGSVTCFGEADOCVACAHYKDP 594
Db      524 SRGVCVTHCNFLNGEPREFAHAEFCFCHPECQMEGTATNGSGSDTCAQCAHFRDGP 583
Qy      595 FCVACRPFNNFTVFWLRVPKYSASHLEPDEGACQPCPINCSTHSC--VDLDDKGCFAEQ 652

```



Db 584 HCVNSCPHGLGAK-----GPIYKPPDAQNECRPHENCTOGCNGPELQDCLGQAEV 635

QY 653 RASPLTSTVSAAVGLLVVLGVFGIILKRRQKIR-KYTWRRLLQETLVEPLTPSGA 711

Db 636 LMSKPHLVIAVVG--LAVIIMILGSGFLYWRGRIRQKRAMRRLYERGESIEPLDPS-E 692

QY 712 MPNQOMKILKETEELRKVKVLSGSAFGTVYKGIWIPDGENVKIPALIVLRENTSPKANK 771

Db 693 KANKVLAIIFKETELRKLVLSGSAFGTVYKGIWIPDGENVKIPALIVLRENTSPKANK 752

QY 772 EILDEAYVAGVSGPYVSRLLGICLTSTVQLVTQLMFPGCLLDHYRENGRLSGQDILNM 831

Db 753 AVTDHMLAVGSLDHAHIVRLGLCFGSSGLQVLTQVLPGLSLDHYKQRETLGQOLLNM 812

QY 832 CMQIAKMSYLEEDVALVHRDLAARVLYKSPNHVKITDFGLARLIDIDETVHADGKVP 891

Db 813 GVQIAKMGVYLEESHSMVHRDLARVLMKSPSOVADVADLLPPDDKQLLHSEAKTP 872

QY 892 IKMMLAESILRRFRTHQSDVSYGVTVLMTFGAKPYDGI-PAREIPPLEKGRRLPQPP 951

Db 873 IKMMLAESIHFGKITHQSDVSYGVTVLMTFGAEFYAGLRALIEPDLKGRRLPQPP 932

QY 952 ICTIDVYIMVYKCMWIDSECRPRRELVSSESRMAKDPQRFVVIQNEDELGPASPLDSTFY 1011

Db 933 ICTIDVYIMVYKCMWIDSECRPRRELVSSESRMAKDPQRFVVIQNEDELGPASPLDSTFY 989

QY 1012 RSLLEDDMGDLVAEEYLVPOGFCPPDPAAGAGMWHHRSSSTRSGGDLTLGEP 1071

Db 990 PSVLTTEL-----QEALPELPEL-----DLDLEA 1015

QY 1072 SEE-----EAPRSPAPASEG-----AGSDVFDGDLGMAKAG 1103

Db 1016 EEEGLATSLGASALSPGTTLRPRGSGQLLSPSSGYPMNQSISGEACLDVAVLGREQF 1075

QY 1104 LQSLPTHPSPLOKRYSEDPVPLPSETDGYV---APL-----TC-----SQPE--- 1144

Db 1076 SRPISLR-PIPRGR-----PASESSEGHVTGSEAEIQEKVSYCRSRSRSPRPRGDS 1127

QY 1145 -YVNOQDVPKPPSPREP-----LPAARPAAGATLERAKTISP-GKNVY--- 1187

Db 1128 AYHQRHSLLPVPLSPGLSEEDGNGYVMDTHLNGASSREGTILSSVGLSGLTSEE 1187

QY 1188 -KVFAFGAVENPEVLTPOGGAPOPHP 1216

Db 1188 EDED-----EEYEMNRKRGRSP-PRPP 1209

RESULT 11

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus

C/Species: amino end of gag protein; env protein fragment; protein-tyrosine kinase

C/Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 11-Jun-1999

C/Accession: B00643 #00643

R/NLSEN, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M

Cell 41, 719-726, 1985

A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and p24

A/Reference number: A00643; MUID:85228222; PMID:2988784

A/Accession: B00643

A/Molecule type: mRNA

A/Residues: 1-698 <NTL>

A/Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750

C/Comment: in Genbank entry CHERB8F, release 109.0, the source is designated as Gallus gallus

C/Comment: This protein is synthesized as a gag-env-erbB protein.

C/Genetics:

A/Gene: gag-env-erbB

C/Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P

F:1-6/Product: gag protein (fragment) #status predicted <GAG>

F:7-59/Product: env protein (fragment) #status predicted <ENV>

F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>

F:194-459/Domain: protein kinase homology <KIN>

F:202-210/Region: protein kinase ATP-binding motif

F:229/Active site: Lys #status predicted

Query Match 25.24; Score 1725.5; DB 1; Length 698;

Best Local Similarity 51.44; Pred. No. 1e-64;

Matches 373; Conservative 76; Mismatches 138; Indels 139; Gaps 20;

QY 578 GPEADQVCAAHYKDPFCVACRPF-----NMFVSTWLVKPVKVSASHLEDEGACQPP 633

Db 60 GP--DHCMKCAHFIDGHCYKACAPVAGENDTL-WMKYA-----DANAQCQLCH 106

QY 634 INCHSVDLDDGCPREQASPLTSTVSAAV-GILLVVLGVFGIILKRRQKIRKYT 692

Db 107 PNCTRGKGGGLEGP---NGSKTPSIAGVVGGLCLVVGIGLYLRR-HYAKRT 162

QY 693 MRLLQETLVEPLTPSGAMPNQOMKILKETEELRKVKVLSGSAFGTVYKGIWIPDGENV 752

Db 163 LRRLQRELVLEPLTPSGEAPNQAHRIKETEFKKVKVLSGSAFGTVYKGIWIPSEKV 222

QY 753 KIVPAIKVLRRENTSPKANKIILDEAYVAGVSGPYVSRLLGICLTSTVQLVTQLMFPGCL 812

Db 223 KIVPAIKVLRRENTSPKANKIILDEAYVAGVSGPYVSRLLGICLTSTVQLVTQLMFPGCL 282

QY 813 LDHRENGRLSGQDILNMCMQIAKMSYLEEDVALVHRDLAARVLYKSPNHVKITDFGL 872

Db 283 LDVIRHKNIGSQVILNMCVQIAKMGVYLEESHSMVHRDLAARVLYKTPQHKITDFGL 342

QY 873 ARLLDIDETVHADGKVPKIMMALESILRRFRTHQSDVSYGVTVLMTFGAKPYDGI 932

Db 343 AKLLGADKEVHAEGCKVPIKMALESILHRIYTHQSDVSYGVTVLMTFGSKPYDGI 402

QY 933 PAREIPPLEKGRRLPQPPCTIDVYIMVYKCMWIDSECRPRRELVSSESRMAKDPQRF 992

Db 403 PASEISVLEKGRRLPQPPCTIDVYIMVYKCMWIDSECRPRRELVSSESRMAKDPQRF 462

QY 993 VVIQ-NEDIGPASPLDSTFYSLLEDDMGDLVAEEYLVPOGFCPPDPAAGAGMWH 1051

Db 463 LVIGQDEHMLPPTPSKFRITLMEEDMEDIDAEYLVPHGFF----- 508

QY 1052 RHSSSTRSGGDLTLGLEPSEBEARSP---APSGAGSDVFDGDLGMAKAGLOS 1106

Db 509 -NSPST-----SRTPLSLSLSTSNMSATNCID-----RNGQG 540

QY 1107 LPHDPSPLORYSEDPVPLPSET-DGVVAPLTPSGPOPEYVNOQDVPKPPRGRPLP 1164

Db 541 HPVREDSFQRYSSDPTGNGFLBESIDGFL-----PAPRYNQ--LMPKRS----- 585

QY 1165 AAPAGATLERAKTISPGKNGVYKVF-----AFGAVENPEVLTPOGGA 1209

Db 586 -----TAMVQNOIYNNISLTAISKLPWDSRYQNSHSTAVDNPEYL----- 625

QY 1210 APOPHPPPAFSAPFDNLYWDD-----DPPE-----RGAPSTKGTPTAENPEY 1254

Db 626 -NTNQSPLAKTVFESSPWIGSNHQLNDPDDYQDLPNETKNGILLKVPAAENPEY 683

QY 1255 LGIDVP 1260

Db 684 LRYAAP 689

RESULT 12

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain

C/Species: avian erythroblastosis virus

C/Date: 18-Apr-1984 #sequence\_revision 18-Apr-1984 #text\_change 11-Jun-1999

C/Accession: A00644; A38022

R/Yamanoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.

Cell 35, 71-78, 1983

A/Title: The erbB gene of avian erythroblastosis virus is a member of the src gene

A/Reference number: A00644; MUID:84026539; PMID:6313229

A/Accession: A00644

A/Molecule type: DNA

A/Residues: 1-604 <YAM>

A/Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678

R:Debutre, B.; Henry, C.; Benalissa, M.; Biserte, G.; Clavierie, J.M.; Saule, S.; Martin, Science 224, 1456-1459, 1984  
 A:Title: Sequencing the erba gene of avian erythroblastosis virus reveals a new type of A:Reference number: A38022; MUID:84223557; PMID:6328658  
 A:Accession: A38022  
 A:Molecule type: DNA  
 A:Residues: 1-28, 'W', '30-139', 'F', '141-145', 'V', '147-152' <DEB>  
 A:Cross-references: GB:K02006  
 C:Genetics:  
 A:Gene: erba  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P  
 F:130-355/Domain: protein kinase homology <KIN>  
 F:138-146/Region: protein kinase ATP-binding motif  
 F:165/Active site: Lys #status predicted

Query Match 24.5% Score 1677; DB 1; Length 604;  
 Best Local Similarity 51.6%; Pred. No. 9 3e-63;  
 Matches 361; Conservative 73; Mismatches 126; Indels 140; Gaps 18;

587 CAHYKDPFCVACRCP---NNFTVSFWLRVKVYASASHLEPDEGACQPCPINCSTHSCVD 642  
 3 CAHFIDGPHCVACAGVAGENDTL---VRKYA-----DANAVCQLCHPNCRCCKG 51  
 643 LDDKCGPAPRASPSTISVAVY-GILLVVLGVVFGILIRKROCKIRKYMRLQETE 701  
 52 PGLGCP---NSKSPSIAGVVGGLVGLVGLIGLYLRR-HYAKRILRLQERE 107  
 702 LVEPLTPSGAMPNQAOMELKTELKRVKVLGSAFGTVYGIWIPDENYKIPAIKVL 761  
 108 LVEPLTPSGEAPNQHRLILKTEFKVKVGLGSAFGITVYGLWIPDEKVKIPAIKEL 167  
 762 RENTSPKANKELIDAVYVWAGSPYVSRLLGICLTSTVOLVQMPYGLLDVHREKRG 821  
 168 REATSPKANKELIDAVYVWASVDNHCVRLLGICLTSTVOLVQMPYGLLDVIREKD 227  
 822 RUGSODLLNMCQIAKMSYLEVDVLRHDLAARVVLKSPNHVITDPGLARLDIDET 881  
 228 NIGSQYLNMCMQIAKMNLYLEERLVRDLAARVVLKTHQKXITDPGLAKLLGADER 287  
 882 EYHADGKVPITKMALESILRRFTHQSDWSYGVTTWELMTFGAKPYDGIIPADEIDLL 941  
 288 EYHAGGKVPITKMALESILHRIYHQSDWSYGVTTWELMTFGSKPYDGIIPASEISVL 347  
 942 EKGERLPQPICTIVYIMVYKWMIDSECRPRELSEFSRMAKRDQRFVIO-NEDL 1000  
 348 EKGERLPQPICTIVYIMVYKWMIDSECRPRELSEFSRMAKRDQRFVIO-NEDL 1000  
 1001 GPASPLDSTFYRSLIEDDMGDLVDAEYLVVPOGFCPPDPAAGAGMWHHRHSSSTRS 1060  
 408 HLPSPDTSKFTYRLTEEDMEDIVDAEYLVPHQGF-----NSPST-- 449  
 1061 GGGDGLTGLPESEEPSPPL-----APSEGAGSVFPGDGLMGAKLQSLPHHDSPL 1115  
 450 -----SKTPLLISLSTASNSATNCTD-----RKGQGHVPREDSFV 485  
 1116 QYSEDPVTPVLPSET--DGYVAPLTCSPQPEYVNGPDVAPQPPSPREGPLPAAPAGATL 1173  
 486 QYSSDPPTGNFLSEIDGFL-----PAPEYVNG--LMPKKPSTAM----- 524  
 1174 ERAKTLSPGKNGVVDVAF-----GGAVENPEVLITQGGAAAPPH 1214  
 555 -----VQNQIYNFISLTAISKLPMSRYQNSHSTAVNDPEYL-----NTN 564  
 1215 PPPAFSPAFDNLVYWDODPPERGAAPSTFGKTPAENPEY 1254  
 565 QSPFAKTVFESSPWYIQSGNHQ-----INDNDY 594

RESULT 13  
 GQPF  
 epidermal growth factor receptor - fruit fly (Drosophila melanogaster)  
 N:Contains: ptyrosine-tyrosine kinase (EC 2.7.1.112) erbb

C:Species: Drosophila melanogaster  
 C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 11-Jun-1999  
 C:Accession: A00640; A38021  
 R:Ulymb, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.  
 Cell 40, 599-607, 1985  
 A:Title: The Drosophila EGF receptor gene homology: conservation of both hormone bit  
 A:Reference number: A00640; MUID:85124611; PMID:2982499  
 A:Accession: A00640  
 A:Molecule type: DNA  
 A:Residues: 1-1330 <LIV>  
 A:Cross-references: EMBL:K03054  
 R:Wadsworth, S.C.; Vincent III, W.S.; Blodgett-Wadsworth, D.  
 Nature 314, 178-180, 1985  
 A:Title: A Drosophila genomic sequence with homology to human epidermal growth fact  
 A:Reference number: A38021; MUID:85137938; PMID:2983232  
 A:Accession: A38021  
 A:Molecule type: DNA  
 A:Residues: 'A', '832-866', 'V', '868-943', 'QTSIVK' <MAD>  
 A:Cross-references: EMBL:X02293; NID:97922; PIDN:CAA26157.1; PID:9292565  
 C:Comment: This sequence is tentative because the introns have not been identified  
 C:Genetics:  
 A:Gene: FlyBase:Egfr  
 A:Cross-references: FlyBase:FBgn0003731  
 A:Map position: 2 57F  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; p  
 F:1-732/Domain: extracellular #status predicted <EXT>  
 F:732-764/Domain: transmembrane #status predicted <TM>  
 F:765-1330/Domain: intracellular #status predicted <INT>  
 F:808-1072/Domain: protein kinase homology <KIN>  
 F:816-824/Region: protein kinase ATP-binding motif  
 F:1122,300,324,363,518,668,695,700/Binding site: carbohydrate (asn) (covalent) #stat  
 F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predic  
 F:843/Active site: Lys #status predicted  
 F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status p

Query Match 23.8% Score 1625.5; DB 1; Length 1330;  
 Best Local Similarity 29.6%; Pred. No. 2.7e-60;  
 Matches 412; Conservative 177; Mismatches 414; Indels 387; Gaps 40;

80 VQGVYLIHNVQVQLORLIRVGTOLF-----EDNYALAVLDNGDPNNTPTVYGASP 134  
 38 ITNYIYGLDILICTSYRLQIRGTLFSLSEKVALFV-----TV 81  
 135 GGLRELQRLTLLKGVILQNPOLCVDTLWMDIFHKNNQALLTIDTRSPACH 194  
 82 SKMYTLEIPDLRVNLGVQGFHNNYVLCNRITIQSEIVSNGTDAYNYDFTAPERECP 141  
 195 CSPMCKGSRGWSSEDCSLTRTVACGCA--RCGKPLPTDCHEQCAAGCTGPRHSDC 252  
 142 CHSCSHG--CWGSGPRKNCQKFSKLTCSPCAGGRGVGPKRRECHLFCAGGCTGPRQKC 200  
 253 LACIHRNHSICGLCPALVTVTDFESPNPNEGRTTGASCYVTCAPNYVLTVDGSC 312  
 201 IACKNFDEAVSEEBPMAKXNPTTVYLETNPEGVAAVAGATVKKCP--CHLRDNGACV 259  
 313 LVCPILNHOEYTAEDGTCRCCKSPCARCYGQVYKANSKFIGTEL-----EPAG 366  
 260 RSCFPQKMDKGE-----CYPCNGPCKTC-----PEVYLHAGNLSFEN 300  
 367 CKKIPLGLAFLESPDG--DPAANTA-----PLQEQLOVFTLEITIGYISAMPDS 418  
 301 CTVIDNIRILDTQTFGQFDVYANYMGPRYIPLDERREVEFTVKEIGYLIESTHQ 360  
 419 LPLSVFONLOVIRGILHNGAV--SLTLQGLGSMGLSLRPLSGLLIHNHTLCV 477  
 361 FRNLSTFRNLLETTHGQLMESFPALAIYKSLYSLEMRNLKQISGSSVVIQHRDLCTV 420  
 478 HTVPMDLFRNPQALLHTANPEDEC----- 504  
 421 SNIRWPAIQEPEQKWNENLRADCGKFLITLISVQNIITHIFAICREKNNHLLGSV 480  
 505 ----- 504

```

Db      481 QGRGLGSMHGSVPYIQELQFQMLHRLMLYIQVINSTODKSNHQLTDCYSPSVPT 540
Qy      505 -----VG 506
Db      541 SLTIERARYAIGSAGLAMELEQITASASMRHSKTLPAEGQVPRWVFLGVCASARAGIA 600
Qy      507 EGLA-----CQGLARGHGWPGPTQCVNCSQFLRGQCEVECRVLOGLREYV---N 556
Db      601 EPLAGRAVCRKCHPPELCTNNGYHEQVCSKLTHTYRRQCECTEC-----PADHYTDEE 654
Qy      557 ARRCLEPCHEPCQPNQSVTCFGEADQVCAAHYK-----DPPF-----CVARCFPN 603
Db      655 QRECFQRPHEC---NG---CTGPGADDCKSCNFKLFDANENGPVYNSTMFCSTKCPLE 708
Qy      604 NPTVSPWLVPKYSASHLEPDEBGACQPCPINCSTCVLDKCGCAEGRASPLTISYA 663
Db      709 -----MRHVNQYVAIGP---YCAASPPRSKITTALD-----NMIFI 744
Qy      664 VVGILLVVLGVVFGI-LIKRQOKIRKYT--MRLLQETELVEPLTPSGAMPNOAMRI 720
Db      745 IIGAVLVPTICLCVTVICRQKAKKETVMTALSGREDESEPLRBNIGANLCKIRI 804
Qy      721 LKETEIRKVKVLSGSGFTVYVIGIMIPDGENYKIPVALKVBRENTSPKANKELDEAYVM 780
Db      805 VVDABELKGGVLMGAFGRVYKGVWVPEGENYKIPVALKELKSTGABESSEFLEAYIM 864
Qy      781 AGVSPVYSRLIGICTSTVOLVTOIMPYGCLLDHVRENGRLSGODLLNMCQIAKMS 840
Db      865 ABEHYNLLKLAVCSSQMLITQMLPGLCLDYVRNRRDKIGSALLNMTQIAKMS 924
Qy      841 YLEDELVYHRLDAARVLYK---SPNHYKITDFGLARLIDIDETEYHADGKVPKIMAL 897
Db      925 YLEERKLVHRDLAARVLYRLAGEDH---DFGLAKLSSDSNHYKAAKGCMPIKMLAL 980
Qy      898 ESILRRPFTQSDWSYGTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITDV 957
Db      981 ECIRNRVFTSKDVAFGVTLTWELLTFQGRPHENIPAKDIPDLIEVGLKLEDFEICSADI 1040
Qy      958 YVIMVYKMWIDSECPREFELVSEFSRMAPDQPRFVVIQNEDLG---PASPLDSTYRSL 1015
Db      1041 YCTLLSCMHLDAAMPTFKQLTVFAEFARDGKRLALIGDKFTLLPA-----YTQD 1093
Qy      1016 EDD---DMGDLVDAEYLYVPOGFCPPDPAPAGAMVHRRSSSTRSGGDLTLGLEPS 1072
Db      1094 EKDLIRKLAFTDGESEALAKPDYLOPKALGPS-----HRTDCT----- 1133
Qy      1073 EEEAP-----RSLAPSEAGSDVFDG---DLGMAKAGLQSLPTHPSPLQRYSEDP 1123
Db      1134 -DEMKLNRYCKDPENKXSTGDDERDSAREVGVGNLR----- 1171
Qy      1124 VPLPSETDGYAPLTCSQPEYVNOQDVRPOPSPREGPLPAPAPAGATLEBAKTLSPGK 1183
Db      1172 LDLPVDEDDYLMF--TCQPGPNNNNNN-----NPNQNMMAVGVAAAYM----- 1214
Qy      1164 NGVAVDVAFAGAVENPEYL---TPQGAAPQPH-----PPAFSP 1221
Db      1215 -----DLIGVPSVSNPEYLLNAQTLGVGESPITQITIGIPWGGPTMEYKVPWPGSEP 1269
Qy      1222 -AFDNLYYWD 1230
Db      1270 TSSDHEYYND 1279

```

RESULT 14  
 S35745  
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus  
 CSpecies: avian erythroblastosis virus  
 CDate: 03-Mar-1994 #sequence\_revision 26-May-1995 #text\_change 28-Feb-1997  
 CAccession: S35745  
 R:Vennistrom, B.  
 Submitted to the EMBL Data Library, March 1993  
 A:Reference number: S35743

```

A:Accession: S35745
A:Molecule type: DNA
A:Residues: 1-544 <VEN>
A:Cross-references: EMBL.X12707
C:Genetics:
A:Gene: erbB
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: Atp; oncogene; phosphotransferase; transforming protein; tyrosine-spec
F:135-400/Domain: protein kinase homology <KIN>
F:143-151/Region: protein kinase Atp-binding motif
F:170/Active site: Lys #status predicted
Query March 23.5% Score 1606; DB 2; Length 544;
Beet Local Similarity 53.9%; Pred. No. 7.38-60;
Matches 344; Conservative 66; Mismatches 122; Indels 106; Gaps 17;
Qy      578 GPEADQVCAHYDPPFCVACRPF-----NFTVSPFLRVKPVASHLEPDEGACQPCP 633
Db      1 GP--DHGKKAHFIDGPHCVACAPAGVLENDTL-VKRYA-----DANVCQLCH 47
Qy      634 INCHSCVDLDDKCPABORASPLTSVAVY-GILLVVLGVVFGILLKROQKIRKYT 692
Db      48 PNCTRGCKGPGLEQCP---NGSKTPSIAAGVVGILLCLVVGIGLGLYLR--HIVKRT 103
Qy      693 MRRLQETELVEPLTPSGAMPNOAMRIKETEIRKVKVLSGSGFTVYVIGIMIPDGENV 752
Db      104 LRLLORELVPEPLTPSGEAPNQHILKETEIRKVKVLSGSGFTVYVIGIMIPDEKV 163
Qy      753 KIPVALKVBRENTSPKANKELDEAYVMAGVSPYRRLGICTSTVOLVTOIMPYGC 812
Db      164 TIPVALKELREATSPKANKELDEAYVMASVDNPHVRLIGICTSTVOLVTOIMPYGC 223
Qy      813 LDHVRNCRGLSGDILLNMCQIAKMSYLEDVRLVHRDLAARVLYKSPNHYKITDFGL 872
Db      224 LDYRERKDNIGSQYLLNMCVQIAKGNVLEERHVRDLAARVLYKTPQHVKITDFGL 283
Qy      873 ARLLIDETEYHADGKVPKIMALESILRRPFTQSDWSYGTVMELMTFGAKPYDGI 932
Db      284 AKQAGABEKYHAGGVPIKMALESILIRITHQSDWSYGTVMELMTFGSKPYDGI 343
Qy      933 PARTIPDLLEKGERLPOPPICITDVIMVYKMWIDSECPREFELVSEFSRMAPDQRF 992
Db      344 PASEISSVLEKGERLPOPPICITDVIMVYKMWIDSECPREFELVSEFSRMAPDQRF 403
Qy      993 VVIO-NEDLGPASLDSTFRSLLEDMDMDLVADEYLYVPOGFCPPDPAPAGAMVHH 1051
Db      404 LVIOGDERMPLPSTDSKFRITLMEEDMEDIVADDEYLVPHQGF----- 449
Qy      1052 RHRSSSTRSGGDLTLGLEPSEEAAPRSP-----APSEAGSDVFPDGLGMGAAGLQ 1106
Db      450 -NSPST-----SRPFLSSLSATSNNSATNCIDRNG----- 480
Qy      1107 LPHTHPSPLQRYSEDPVPLPSETDGYAPLTCSQPEYVNOQDVRPOPSPREGPLPAA 1166
Db      481 --H-----PVREDDGL-----PAPEYVNO--LMEKPKSTAMVQNOIY 513
Qy      1167 RPAGAT-LEBAKTLSPKNGVAVDVFAGAVENPEYL 1203
Db      514 NYISLRAISKLPIDSKYQ-----SHSTAVVNPEYL 544

```

RESULT 15  
 S00727  
 kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi  
 CSpecies: avian erythroblastosis virus  
 CDate: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 28-Feb-1997  
 CAccession: S00727  
 R:Scottling, P.; Vennistrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.  
 Oncogene Res. 1, 265-278, 1987  
 A>Title: Common site of mutation in the erbB gene of avian erythroblastosis virus r  
 A:Reference number: S00727; MUID:88217326; PMID:2897102  
 A:Accession: S00727  
 A:Molecule type: DNA

A:Residues: 1-545 <SC0>  
A:Cross-references: EMBL:X06943

C:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; phosphotransferase

F:135-400/Domain: protein kinase homology <KIN>  
F:143-151/Region: protein kinase ATP-binding motif

Query Match 23.4%; Score 1599; DB 2; Length 545;  
Best Local Similarity 53.9%; Pred. No. 1,4e-59;  
Matches 344; Conservative 65; Mismatches 123; Indels 106; Gaps 17;

```
QY 578 GPEADQCVACAHYKDPFCVACPF---NNFTSPFLRVPKVASHLPEDEGACQPCP 633
DB 1 GP--DHQMKCAHFDGPHCVKACPAVLGENDTL-VKXA-----DANAVQQLCH 47
QY 634 INCTHSCVDLDKGCAPBQAPSLTIVSAV-GILLVVVLGVFGILIKRQOKIRKKT 692
DB 48 PNCTRGCKGPGLEGCP---NGSKTPSIAGVVGILCLVVGIGLGLYLRR-HIVRKRT 103
QY 693 MRRLQETELVEPLTPSGAMPNOAKRILKETELRKVYLGSGAFGTYYKGIWIPGENV 752
DB 104 LRRLQERELVEPLTPSGAPNOAHRLKETEFKVKYLGGAFTYKGLMDEGEKV 163
QY 753 KIPVAIKVRENTSPANKELIDEAYVMAVGSPVYSRLIGICTSTVQLTQMPYGCL 812
DB 164 TIPVAIKELREATSPANKELIDEAYVMAVDNPHVCRLLGICTSTVQLTQMPYGCL 223
QY 813 LDHVRNRGLSGODLLNMCQIAKMSYLEPRLVHRDLAARNVLYKSPNHKITTDEL 872
DB 224 LDYIRHKDNIGSQYLLNMCVQIAKNNYLEERHLVHRDLAARNVLYKTPQDKITDFGL 283
QY 873 ARLLDIDETEVHADGKVPDKMALESILRRFTHSDVMSYGVTVWELMTFGAKPYDGI 932
DB 284 AKQLGADKEKYEHAEGKVPDKMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI 343
QY 933 PAREIPDLLEKGRLLPQPICTIDVYMIWYKCMIDSECRPRRELVSFSPMRADPQRF 992
DB 344 PASEISSVLEKGERLPQPICTIDVYMIWYKCMISDADSRPKFRELIAEFSKMRADPPRY 403
QY 993 VVIQ-NEDLGPAAPLDSTFYRSLLEDMDMDLVDABEYLVPQGFPCPDPAAGAGMVH 1051
DB 404 LVIQSDPRMHLPSPTSKFYRLTMEEDMEDIVADBYLVPHQGF----- 449
QY 1052 RHRSSSTRSGGGLTLGLEPSEBEAPRSL---APSEGAGSDVDFDGLGMGAAGLQS 1106
DB 450 --NSPST-----SRTPLSLSLSATSNNSATNCIDRNG----- 480
QY 1107 LPTHDPSPLORYSEDPVLPSETDGVVAPLTCSPQEVYNOQDVPPQPPSPREGPLPA 1166
DB 481 ---H-----PYREDGFL-----PAPEYVNC--LMPKRPSTAMVONQIY 513
QY 1167 RPAGAT-LEBAKTLSPKRGVVKDVFAGGAVENPEYL 1203
DB 514 NYISLTAISKLPMDSRYN-----SHSTAVDNPEYL 544
```

Search completed: July 22, 2003, 09:10:21  
Job time : 32.0157 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:12:49 ; Search time 10.2793 Seconds

(Without alignments)  
5088.033 Million cell updates/sec

Title: SEQ4-579-593-14

Perfect score: 6841

Sequence: 1 MEUAALCRMGCLLALLPPGA.....TEKGTPTAENPEYLGDVPEV 1261

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6621	96.8	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5833	85.3	1257	1 ERB2_RAT	P06494 rattus norv
3	5817.5	85.0	1254	1 ERB2_MESAU	P06553 mesocricetu
4	3108	45.4	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3080	45.0	1210	1 EGFR_MOUSE	P01279 mus musculu
6	2945.5	43.1	1308	1 ERB4_HUMAN	O15303 homo sapien
7	2929	42.8	1308	1 ERB4_RAT	P02986 rattus norv
8	2643.5	38.6	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2387.5	34.9	1342	1 ERB3_HUMAN	P21860 homo sapien
10	2316.5	33.9	1339	1 ERB3_RAT	P062799 rattus norv
11	1960	28.7	1426	1 EGFR_DROME	P00534 avian leuko
12	1708.5	25.0	634	1 ERBB_AYER	P00535 avian eryth
13	1677	24.5	604	1 ERBB_AYER	P11273 avian eryth
14	1589	23.2	540	1 EGFR_CHICK	P13387 gallus gall
15	1584	22.7	703	1 EGFR_MOUSE	P24338 caenorhabdi
16	1305	19.1	1323	1 LIT23_CAEEL	P02438 mus musculu
17	1142.5	16.7	245	1 ERB2_MOUSE	O02466 branchiosto
18	730	10.7	1363	1 ILPR_BRLA	O09114 mus musculu
19	710	10.4	1300	1 IRR_MOUSE	P14616 homo sapien
20	695	10.2	1297	1 IRR_HUMAN	P06223 homo sapien
21	694	10.1	1382	1 IRR_HUMAN	P15127 rattus norv
22	694	10.1	1383	1 IRR_RAT	P025197 hydra atten
23	693	10.1	1477	1 HTK_HYDAT	P025197 mus musculu
24	691.5	10.1	1372	1 IRR_MOUSE	O25410 lymnaea ste
25	691	10.1	1607	1 MIPR_LYMT	P14617 cavia porce
26	688.5	10.1	1300	1 IRR_CAVPO	P08069 homo sapien
27	649	9.5	1367	1 IRR_HUMAN	P060751 mus musculu
28	629	9.2	1373	1 IRR_MOUSE	P24062 rattus norv
29	625.5	9.1	1370	1 IRR_RAT	P093105 aedes aegypt
30	620	9.1	1390	1 IRR_AEDAE	P092028 drosophila
31	614	9.0	2146	1 INSR_DROME	P54760 homo sapien
32	597	8.7	987	1 EPB4_HUMAN	P03145 mus musculu
33	594.5	8.7	977	1 EPB4_MOUSE	

34	591.5	8.6	984	1 EPB1_RAT	P09759 rattus norv
35	588.5	8.6	976	1 EPB2_HUMAN	P29317 homo sapien
36	586.5	8.6	984	1 EPB1_CHICK	P07494 gallus gall
37	585.5	8.6	984	1 EPB1_HUMAN	P54762 homo sapien
38	583	8.5	902	1 EPB3_XENLA	O91736 xenopus lae
39	581.5	8.5	985	1 EPB4_XENLA	O91571 xenopus lae
40	577	8.4	987	1 EPB4_MOUSE	P54761 mus musculu
41	574.5	8.4	1053	1 FAK1_MOUSE	O00944 gallus gall
42	574	8.4	1114	1 RET_HUMAN	P07949 homo sapien
43	569	8.3	1068	1 FAK1_XENLA	O91738 xenopus lae
44	566	8.3	757	1 HTI6_HYDAT	P53356 hydra atten
45	563	8.2	1052	1 FAK1_MOUSE	P34152 mus musculu

## ALIGNMENTS

RESULT 1  
ID ERB2\_HUMAN STANDARD; PRT; 1255 AA.  
AC P04626;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell  
DE surface receptor HER2) (MLN 19).  
GN ERB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86118663; PubMed=3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,  
RA Saito T., Toyoshima K.;  
RT "Similarity of protein encoded by the human c-erb-B-2 gene to  
RT epidermal growth factor receptor.";  
RL Nature 319:230-234 (1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86070181; PubMed=2999974;  
RA Gauseens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,  
RA McGarrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,  
RA Francke U., Levinson A., Ullrich A.;  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor  
RT shares chromosomal location with neu oncogene.";  
RL Science 230:1132-1139 (1985).  
RN [3]  
RP SEQUENCE OF 737-1031 FROM N.A.  
RX MEDLINE=8616729; PubMed=2995967;  
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.,  
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the  
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a  
RT human salivary gland adenocarcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985).  
RN [4]  
RP VARIANTS VAL-654 AND VAL-655.  
RX MEDLINE=93194196; PubMed=8095488;  
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;  
RT "Characterization of a new allele of the human ERBB2 gene by allele-  
RT specific competition hybridization.";  
RL Genomics 15:426-429 (1993).  
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
CC ALPHA AND AMPHIREGULIN.  
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC (POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- P1M: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY).  
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN  
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ch/announce/>  
 CC or send an email to [license@ebi.ch](mailto:license@ebi.ch)).  
 CC -----  
 DR EMBL; M11767; AAA5808.1; JOINED.  
 DR EMBL; M11761; AAA5808.1; JOINED.  
 DR EMBL; M11762; AAA5808.1; JOINED.  
 DR EMBL; M11763; AAA5808.1; JOINED.  
 DR EMBL; M11764; AAA5808.1; JOINED.  
 DR EMBL; M11765; AAA5808.1; JOINED.  
 DR EMBL; M11766; AAA5808.1; JOINED.  
 DR EMBL; M11730; AAA5808.1; JOINED.  
 DR EMBL; M12036; AAA5808.1; JOINED.  
 DR EMBL; X03363; CAA27060.1; JOINED.  
 DR PIR; A25491; A25491.  
 DR PIR; A24571; A24571.  
 DR HSSP; F11362; 1FGK.  
 DR Genew; HGNC:3430; ERBB2.  
 DR MIM; 164870; ERBB2.  
 DR Interpro; IPR000494; EGFR\_L\_domain.  
 DR Interpro; IPR000719; Euk\_kinase.  
 DR Interpro; IPR002174; Furin-like.  
 DR Interpro; IPR001245; Tyr\_kinase.  
 DR Interpro; IPR004019; YLP\_motif.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR Pfam; PF02757; YLP; 2.  
 DR Prodom; P0000001; Euk\_kinase; 1.  
 DR SMART; SM00261; Fu; 3.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Polymorphism.  
 KW SIGNAL 1 21  
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 653 675 POTENTIAL.  
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 720 987 PROTEIN KINASE.  
 FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
 FT BINDING 753 753 ATP (BY SIMILARITY).  
 FT ACT\_SITE 845 845 BY SIMILARITY.  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 220 227 BY SIMILARITY.  
 FT DISULFID 224 235 BY SIMILARITY.  
 FT DISULFID 236 244 BY SIMILARITY.  
 FT DISULFID 240 252 BY SIMILARITY.  
 FT DISULFID 255 264 BY SIMILARITY.  
 FT DISULFID 268 295 BY SIMILARITY.  
 FT DISULFID 299 311 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 511 520 BY SIMILARITY.  
 FT DISULFID 515 528 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 567 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARIANT 654 654 I -> V.  
 FT VARIANT 654 654 /FTID=VAR\_004077.  
 FT VARIANT 655 655 I -> V.  
 FT CONFLICT 1170 1170 /FTID=VAR\_004078.  
 FT SEQUENCE 1255 AA; 137909 MM; 39E9DFDA04DF962 CRC64;  
 SQ  
 Query Match 96.8%; Score 6621; DB 1; Length 1255;  
 Best Local Similarity 96.8%; Pred. No. 0; Mismatches 19; Indels 14; Gaps 2;  
 Matches 1225; Conservativity 7;  
 QY 1 MELAALCRWGLTALPPGAASVQVCTGDMKRLPASPETHLMLRHLHYQGQVQGNL 60  
 DB 1 MELAALCRWGLTALPPGAASVQVCTGDMKRLPASPETHLMLRHLHYQGQVQGNL 60  
 QY 61 ELTYLPFNASLSTLOIEVQGVYLAHQVQVPLQRLRYRGSTOLFEDNYALAVLDNG 120  
 DB 61 ELTYLPFNASLSTLOIEVQGVYLAHQVQVPLQRLRYRGSTOLFEDNYALAVLDNG 120  
 QY 121 DPLNNTPTVPGASPGGRLQRLSTLEIKGVLQGNPOLCYQDTILMKDIFHKNNOLA 180  
 DB 121 DPLNNTPTVPGASPGGRLQRLSTLEIKGVLQGNPOLCYQDTILMKDIFHKNNOLA 180  
 QY 121 DPLNNTPTVPGASPGGRLQRLSTLEIKGVLQGNPOLCYQDTILMKDIFHKNNOLA 180  
 DB 121 DPLNNTPTVPGASPGGRLQRLSTLEIKGVLQGNPOLCYQDTILMKDIFHKNNOLA 180  
 QY 181 LTLIDNRSBACHPCSPCKGRCWESSEDCSLTRVYACGACACGKPLPDCHEQC 240  
 DB 181 LTLIDNRSBACHPCSPCKGRCWESSEDCSLTRVYACGACACGKPLPDCHEQC 240  
 QY 181 LTLIDNRSBACHPCSPCKGRCWESSEDCSLTRVYACGACACGKPLPDCHEQC 240  
 DB 181 LTLIDNRSBACHPCSPCKGRCWESSEDCSLTRVYACGACACGKPLPDCHEQC 240  
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDPESMNPGRVYFGASCYACP 300  
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDPESMNPGRVYFGASCYACP 300  
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDPESMNPGRVYFGASCYACP 300  
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDPESMNPGRVYFGASCYACP 300  
 QY 301 YNLTSTVSGCTIVCPAHNQEYTAEDGTQCEKSCRPACVYGLGMQYIKANSKFIGIT 360  
 DB 301 YNLTSTVSGCTIVCPAHNQEYTAEDGTQCEKSCRPACVYGLGMQYIKANSKFIGIT 360  
 QY 301 YNLTSTVSGCTIVCPAHNQEYTAEDGTQCEKSCRPACVYGLGMQYIKANSKFIGIT 360  
 DB 301 YNLTSTVSGCTIVCPAHNQEYTAEDGTQCEKSCRPACVYGLGMQYIKANSKFIGIT 360  
 QY 361 ELEPAGCKTIFGSLAFIPESFGDDPASNTPAPLOPQLOVETLEITGYIYIAMPDLSL 420  
 DB 361 ELEPAGCKTIFGSLAFIPESFGDDPASNTPAPLOPQLOVETLEITGYIYIAMPDLSL 420  
 QY 361 IOEPAGCKTIFGSLAFIPESFGDDPASNTPAPLOPQLOVETLEITGYIYIAMPDLSL 420  
 DB 361 IOEPAGCKTIFGSLAFIPESFGDDPASNTPAPLOPQLOVETLEITGYIYIAMPDLSL 420  
 QY 421 DLSVFOQLQVIRGRILHNGAYSLTLOGLGSMGLSLRELGSGLALIHNTLCEVHTV 480  
 DB 421 DLSVFOQLQVIRGRILHNGAYSLTLOGLGSMGLSLRELGSGLALIHNTLCEVHTV 480  
 QY 481 PMDLPFNPHQALLHTANRDEDCVAGGLACHQLCARGHGCGFPQTCVNCQFPLRQEC 540  
 DB 481 PMDLPFNPHQALLHTANRDEDCVAGGLACHQLCARGHGCGFPQTCVNCQFPLRQEC 540  
 QY 481 PMDLPFNPHQALLHTANRDEDCVAGGLACHQLCARGHGCGFPQTCVNCQFPLRQEC 540  
 DB 481 PMDLPFNPHQALLHTANRDEDCVAGGLACHQLCARGHGCGFPQTCVNCQFPLRQEC 540  
 QY 541 VEEPRVYQGLPREVNARHCLPCHPEGQPNQSVTCGPPADCCVACAHKDPFCVARC 600  
 DB 541 VEEPRVYQGLPREVNARHCLPCHPEGQPNQSVTCGPPADCCVACAHKDPFCVARC 600  
 QY 541 VEEPRVYQGLPREVNARHCLPCHPEGQPNQSVTCGPPADCCVACAHKDPFCVARC 600  
 DB 541 VEEPRVYQGLPREVNARHCLPCHPEGQPNQSVTCGPPADCCVACAHKDPFCVARC 600  
 QY 601 PFNNFTVFWLRYPKVSASHLE---PDEGACQPCPINCTHSCVDLDDKGAQPAQASP 656  
 DB 601 PSQ-----VKPDLSTYMPIMKPFDEGACQPCPINCTHSCVDLDDKGAQPAQASP 650  
 QY 657 LTSIVSANGVGLLVVYGVVGLILKRRQKIRKTYRRLLOSTELVEPLTPSGAMPNCA 716  
 DB 651 LTSIISAVGILLVVVGVVGLILKRRQKIRKTYRRLLOSTELVEPLTPSGAMPNCA 710

```

QY 717 QMILKETEELKRVKVLGSGAFGTIVKGIWIPDGENVKI PVAIKVLRENTS PKANKELIDE 776
DB 711 QMRLKETEELKRVKVLGSGAFGTIVKGIWIPDGENVKI PVAIKVLRENTS PKANKELIDE 770
QY 777 AYVWAGVSPYVSSLLGICLTSTVQLVTOQMPFGCLLDHRENRGRIGSODLLNMCQIA 836
DB 771 AYVWAGVSPYVSSLLGICLTSTVQLVTOQMPFGCLLDHRENRGRIGSODLLNMCQIA 830
QY 837 KGNVSYLDEVLVHEDLAAARNVLPKPNVKITDGLARLIDIDETEHADGKVPKMA 896
DB 831 KGNVSYLDEVLVHEDLAAARNVLPKPNVKITDGLARLIDIDETEHADGKVPKMA 890
QY 897 LESILRRRFTHQSDVMSYGTIVELMTFGAKPIDGIPAREIPDLLEKGERLPQPICTID 956
DB 891 LESILRRRFTHQSDVMSYGTIVELMTFGAKPIDGIPAREIPDLLEKGERLPQPICTID 950
QY 957 VYIMVWCMVIDSCRFRELVSEFSRMARDQRFVIVIONEDLGAPSPIDSTFFYSILE 1016
DB 951 VYIMVWCMVIDSCRFRELVSEFSRMARDQRFVIVIONEDLGAPSPIDSTFFYSILE 1010
QY 1017 DDMWGDVDAEYLVPOQGFPCDPAPAGAGVYHRRSSSTRSGGDLTLGLEPSEEA 1076
DB 1011 DDMWGDVDAEYLVPOQGFPCDPAPAGAGVYHRRSSSTRSGGDLTLGLEPSEEA 1070
QY 1077 PRSPLASBAGSDVFDGDLGMGAAGKLSLPHDPSPLQRYSEDDPVLPSTTDGVAP 1136
DB 1071 PRSPLASBAGSDVFDGDLGMGAAGKLSLPHDPSPLQRYSEDDPVLPSTTDGVAP 1130
QY 1137 LTCSPQPEYVNOQDVPFPPSPREGPLPAARPAATLERAKTSPGNKVGVKVFAGGA 1196
DB 1131 LTCSPQPEYVNOQDVPFPPSPREGPLPAARPAATLERAKTSPGNKVGVKVFAGGA 1190
QY 1197 VENPEYLTPOGGAAPQHPAPPATSPANDLTYWDOPPEFGAPSTFKGTIANPEYIG 1256
DB 1191 VENPEYLTPOGGAAPQHPAPPATSPANDLTYWDOPPEFGAPSTFKGTIANPEYIG 1250
QY 1257 LDVPV 1261
DB 1251 LDVPV 1255

RESULT 2
ERRB2 RAT STANDARD; PRT; 1257 AA.
AC P06454;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbb-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor
DE receptor-related protein).
GN ERBB2 OR NEU
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RA MEDLINE=6118662; PubMed=3945311;
RA Bergmann C.I., Hung M.-C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RA MEDLINE=9122560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).

```



FT DISULFID 517 530 BY SIMILARITY.  
 FT DISULFID 533 542 BY SIMILARITY.  
 FT DISULFID 546 562 BY SIMILARITY.  
 FT DISULFID 565 578 BY SIMILARITY.  
 FT DISULFID 569 586 BY SIMILARITY.  
 FT DISULFID 589 598 BY SIMILARITY.  
 FT DISULFID 602 625 BY SIMILARITY.  
 FT DISULFID 628 636 BY SIMILARITY.  
 FT DISULFID 632 644 BY SIMILARITY.  
 FT MOD\_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).  
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 85.3%; Score 5833; DB 1; Length 1257;  
 Best Local Similarity 85.4%; Pred. No. 5,4e-305;  
 Matches 1082; Conservative 53; Mismatches 116; Indels 16; Gaps 4;

QY 1 MELALCRWGLLALPPGAASVCTGTDMKRLPASPEHLDMLSHLYOGCQVQGNL 60  
 DB 1 MELAAWCGFLALALPPGIAGTCTGTDMKRLPASPEHLDMLSHLYOGCQVQGNL 60  
 QY 61 ELTYLPTNASLSPLODIOEVGYVLIANOVROYPLORLRIVRGSTOLFEDNYALAVLDNG 120  
 DB 61 ELTYVPAVASLSPLDIOEVGYVLIANOVKRVPLORLRIVRGSTOLFEDNYALAVLDNR 120  
 QY 121 DPLNNTTPEVT-GASPGSLRELQSLTEILKGVYLIQNPOLCYQDTILMKDIFHKNNOL 179  
 DB 121 DPQDVAASVTPGRTEGRELQSLTEILKGVYLIQNPOLCYQDVMVLMKDVFRKNNOL 180  
 QY 180 ALTLIDTRSPACHPSCPMCKGSGKSCGSSSEDCSLRTVACGACRCKGLPPTDCHQ 239  
 DB 181 APVIDITNRSRACPCACAPACKDNHCWGESPECCQLVGTICTSCARCKGLPPTDCHQ 240  
 QY 240 CAAGCTGPKHSDCLACHFNHSGICEIHPALVYNTDFESMNPREGVTFPGASCTVAC 299  
 DB 241 CAAGCTGPKHSDCLACHFNHSGICEIHPALVYNTDFESMNPREGVTFPGASCTVTC 300  
 QY 300 PYNVISTVSGCTLVCLAHQEBVTADGTQCEKSKCAQAVCYIGLMQYIKANSKRTGI 359  
 DB 301 PYNVISTVSGCTLVCPNNQEBVTADGTQCEKSKCAQAVCYIGLMQYIKANSKRTGI 360  
 QY 360 TELEPAGCKXIFGSLAFIPESPDGPASNTAPLOPELOQVETLEIETGYIYISAMPDSL 419  
 DB 361 NVQEPDCKKIFGSLAFIPESPDGPASNTAPLOPELOQVETLEIETGYIYISAMPDSL 420  
 QY 420 PDLVSFQNLQVIRGILHNGAVSLTQGLGSMWGLSRLSGGLAIHNNHLCFVHT 479  
 DB 421 RDLVSFQNLRIIRGILHNGAVSLTQGLGSHLGRSLRSLSGGLAIHNNHLCFVHT 480  
 QY 480 VPMQGLFENPHOALLHTANRPEDE-CVGEGLACHQCLARGCMGPGPTOCNCSQPLRG 538  
 DB 481 VPMQGLFENPHOALLHTANRPEDE-CVGEGLACHQCLARGCMGPGPTOCNCSQPLRG 540  
 QY 539 ECVEECRVLQGLPREYVNAHRCIHPCEPCQPNQSVTGFGEADQCVACAYKDPFCVA 598  
 DB 541 ECVEECRVWKGIPREYVNAHRCIHPCEPCQPNQSVTGFGEADQCVACAYKDPFCVA 600  
 QY 599 RCPFNFTVSMFLARPKXSASHLE---PDEGACQCPINICTHSVDLDDKGPAPQRA 654  
 DB 601 RCPFG-----VPEDLSTMPWKYPDEBGICQCPINICTHSVDLDDKGPAPQRA 650  
 QY 655 SPLTIVSAVVGILLVVLGVFGILLIRKROOKIRKXTRMLLOTELVEBLVETSGAMPN 714  
 DB 651 SPLTIVSAVVGILLVVLGVFGILLIRKROOKIRKXTRMLLOTELVEBLVETSGAMPN 710  
 QY 715 QAQMKILKETELRKXVUGSAGFVYVIGIWIPOGENVKIIVALKVLRNTSPKANKEIL 774

DB 711 QAQMKILKETELRKXVUGSAGFVYVIGIWIPOGENVKIIVALKVLRNTSPKANKEIL 770  
 QY 775 DEAYVAGVSGPVYSRLGICITSTVQVLTQMPYGLLDHVRNKGSLGQDLNMCQ 834  
 DB 771 DEAYVAGVSGPVYSRLGICITSTVQVLTQMPYGLLDHVRNKGSLGQDLNMCQ 830  
 QY 835 IAKGMSYLEDVLRDLAARNVLYKSPHNVITPFGRLALDIDETEXHAQGVPIK 894  
 DB 831 IAKGMSYLEDVLRDLAARNVLYKSPHNVITPFGRLALDIDETEXHAQGVPIK 890  
 QY 895 MALESILRRRPTHQSDVMSYGVTELMFGAKPYDGIIPAREIPDLEKGERLPQPICT 954  
 DB 891 MALESILRRRPTHQSDVMSYGVTELMFGAKPYDGIIPAREIPDLEKGERLPQPICT 950  
 QY 955 IDVYIMVYCMWIDSCRRPREFVSEFRMARDPQRFVIONEDGPSPLDSTYRSL 1014  
 DB 951 IDVYIMVYCMWIDSCRRPREFVSEFRMARDPQRFVIONEDGPSPLDSTYRSL 1010  
 QY 1015 LEDDDMGDLVDAEYLVPOQGFPCDPAPAGAMVHRRSSSTRSGGDLTLGLEPSEE 1074  
 DB 1011 LEDDDMGDLVDAEYLVPOQGFPCDPAPAGAMVHRRSSSTRSGGDLTLGLEPSEE 1070  
 QY 1075 EAPRSLAPSEAGSDVPEPDIQKGAAGLQSLPHDSEPLQYSEDPIVPLPSETDGY 1134  
 DB 1071 EAPRSLAPSEAGSDVPEPDIQKGAAGLQSLPHDSEPLQYSEDPIVPLPSETDGY 1130  
 QY 1135 APLTCSPOPEYVNPQVRPOPSPREGPLPAARPAGATLERAKTSPGKNGVYKQVFAFG 1194  
 DB 1131 APLTCSPOPEYVNPQVRPOPSPREGPLPAARPAGATLERAKTSPGKNGVYKQVFAFG 1190  
 QY 1195 GAVENBEYLTPOGGAAPQHPAPAFSPAFDNLYWDDPDERGAPPESTKGTPTANPEY 1254  
 DB 1191 GAVENBEYLTPOGGAAPQHPAPAFSPAFDNLYWDDPDERGAPPESTKGTPTANPEY 1250  
 QY 1255 LGIDVPV 1261  
 DB 1251 LGIDVPV 1257

RESULT 3  
 ERB2\_MESAU STANDARD; PRT; 1254 AA.  
 ID ERB2\_MESAU  
 AC Q60553;  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
 GN ERBB2 OR NEU.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 OX NCBI\_Taxid=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Neve;  
 RX MEDLINE=94193007; PubMed=7908275;  
 RA Nakamura T., Tahijima T., Ishizaka Y., Nagao M., Arai M.,  
 RA Yamazaki Y., Ishikawa T.;  
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene."  
 RL Gene 140:251-255(1994).  
 CC -|- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).  
 CC -|- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -|- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -|- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -|- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE

RESIDUES.  
- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
-----  
CC EMBL: D16295; BAA03801.1;  
CC HSSP: P11352; 1FCX.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR InterPro: IPR004019; YLP\_mocif.  
DR Pfam: PF00757; Furin-like; 1.  
DR Pfam: PF01030; Recep\_L\_domain; 2.  
DR Pfam: PF02757; YLP; 2.  
DR ProDom: PD000001; Euk\_pkinase; 1.  
DR SMART: SM00261; FU; 3.  
DR SMART: SM00219; Tyrc; 1.  
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE: PS00109; PROTEIN KINASE\_TYR; 1.  
DR PROSITE: PS50011; PROTEIN KINASE\_DOM; 1.  
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
KW Proto-oncogene; Disease mutation.  
FT CHAIN 1 21 POTENTIAL.  
FT SIGNAL 1 21  
FT DOMAIN 22 1254 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
FT TRANSMEM 653 675 EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 158 368 CYS-RICH.  
FT DOMAIN 472 644 CYS-RICH.  
FT NP\_BIND 720 987 PROTEIN KINASE  
FT BINDING 726 734 ATP (BY SIMILARITY).  
FT ACT\_SITE 753 753 ATP (BY SIMILARITY).  
FT ACT\_SITE 845 845 BY SIMILARITY.  
FT DISULFID 195 204 BY SIMILARITY.  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 236 244 BY SIMILARITY.  
FT DISULFID 240 252 BY SIMILARITY.  
FT DISULFID 255 264 BY SIMILARITY.  
FT DISULFID 268 295 BY SIMILARITY.  
FT DISULFID 299 311 BY SIMILARITY.  
FT DISULFID 315 331 BY SIMILARITY.  
FT DISULFID 334 338 BY SIMILARITY.  
FT DISULFID 511 520 BY SIMILARITY.  
FT DISULFID 515 528 BY SIMILARITY.  
FT DISULFID 531 540 BY SIMILARITY.  
FT DISULFID 544 560 BY SIMILARITY.  
FT DISULFID 563 576 BY SIMILARITY.  
FT DISULFID 584 584 BY SIMILARITY.  
FT DISULFID 587 596 BY SIMILARITY.  
FT DISULFID 600 623 BY SIMILARITY.  
FT DISULFID 626 634 BY SIMILARITY.  
FT DISULFID 630 642 BY SIMILARITY.  
FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT MOD\_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 125 125 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).  
FT CARBOHYD 658 658 N-LINKED (GLCNAC...) (POTENTIAL).  
FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).  
FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).  
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.0%; Score 5817.5; DB 1; Length 1254;  
Best Local Similarity 85.0%; Pred. No. 3,66-304;  
Matches 1075; Conservative 62; Mismatches 113; Indels 15; Gaps 3;

QY	1	MELALCRWGLLALLPPGAASVCTGTDMKRLRPLASPEHLDMLHLYOGCOVQNTL	60
DB	1	MELAAQGWGLLALLPPGAASVCTGTDMKRLRPLASPEHLDMLHLYOGCOVQNTL	60
QY	61	ELTLPNNAISFLDIOEVQGYLTHANOVRQVPLRLRVRSTOLPEDNYALAVDNG	120
DB	61	ELTLPNNAISFLDIOEVQGYLTHANOVRQVPLRLRVRSTOLPEDNYALAVDNG	120
QY	121	DLPLNTTPTVGASPGGLREQLRLSLTEILKGVLIQNPOLCYODTILMKDIPIKNNOLA	180
DB	121	DLPLNTTPTVGASPGGLREQLRLSLTEILKGVLIQNPOLCYODTILMKDIPIKNNOLA	180
QY	181	LTLLDTNRSRACHCSFCKSGSRGWSSSDCCSLTITVACGACRCKPLPTDCCHQC	240
DB	181	PVDIDTNRSRACPCPCACCKDNHGMGASPEDCCOTLTITIAVPAARARLPTDCCHQC	240
QY	241	AACTGPKHSDCLAFHNSGICELCPALVTNTDTPESMPREGRTFGASCYACP	300
DB	241	AACTGPKHSDCLAFHNSGICELCPALVTNTDTPESMPREGRTFGASCYACP	300
QY	301	YNYLSTDVSGCTLVCPILHNOEVTAEQDQCEKSKPCARVYCYGLMOTYKANSKFIGIT	360
DB	301	YNYLSTDVSGCTLVCPILHNOEVTAEQDQCEKSKPCARVYCYGLMOTYKANSKFIGIT	360
QY	361	ELFENAGCKTFGSLAPLPSFDPDGPANTAPLOEQVETLEITGYLYISAWPDSL	420
DB	361	IOEPAGCKTFGSLAPLPSFDPDGPANTAPLOEQVETLEITGYLYISAWPDSL	420
QY	421	DLSTFQNLQVIRGRILHNGAVSLTLOGLISWGLRSRLRELSGLAIHNNTHLCFHTV	480
DB	421	DLSTFQNLQVIRGRILHNGAVSLTLOGLISWGLRSRLRELSGLAIHNNTHLCFHTV	480
QY	481	PMQDLFNNPQALHTNRPEDBCVSGELACHQICARHGMCPGPTQCVNCSQFLRQEC	540
DB	481	PMQDLFNNPQALHTNRPEDBCVSGELACHQICARHGMCPGPTQCVNCSQFLRQEC	540
QY	541	VEEGRVYQGLPREVYVNRHCLPCHPEQOPONGSVTCGPRAADCVAAHKKDPGCVARC	600
DB	541	VKEGRVYQGLPREVYVNRHCLPCHPEQOPONGSVTCGPRAADCVAAHKKDPGCVARC	600
QY	601	PFNNFTVSWLVRVPKVSASHLE---PDEBGAQCPCEINCTHSCVDLDDKCPAEPQASP	656
DB	601	PSG-----VKPDLSTYMPIMKYVDEBGMQPCINCTHSCVDLDDKCPAEPQASP	650
QY	657	LTSIVSAYVGLLVVYGVFGILIRKROKIRKTYMRLLQETELVEPLTPSGAMPNOA	716
DB	651	ATSLIATVAVGLLVVYGVFGILIRKROKIRKTYMRLLQETELVEPLTPSGAMPNOA	710
QY	717	QMRILKETELRKVYVGSAGFVYKGIWIPDENYKIPAIKVLRENTSPKANKELIDE	776
DB	711	QMRILKETELRKVYVGSAGFVYKGIWIPDENYKIPAIKVLRENTSPKANKELIDE	770
QY	777	AYWAGVGSPPYRBLGICLTSTVQVLTQMPYGLLDHVENRGRIGSODLLMCMQIA	836
DB	771	AYWAGVGSPPYRBLGICLTSTVQVLTQMPYGLLDHVENRGRIGSODLLMCMQIA	830
QY	837	KGMSYLEDVAVLRDIAARVAVLKSPHNVITDPGLARLIDIDETETHAGGVPIKMA	896
DB	831	KGMSYLEDVAVLRDIAARVAVLKSPHNVITDPGLARLIDIDETETHAGGVPIKMA	890
QY	897	LESILRRFTHSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID	956
DB	891	LESILRRFTHSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTID	950
QY	957	VYVIMVYKMWIDSECPREPRFELVSEFSMAADQQRVVVIONEDLGASPLDSTFYRSLLE	1016
DB	951	VYVIMVYKMWIDSECPREPRFELVSEFSMAADQQRVVVIONEDLGASPLDSTFYRSLLE	1010

QY 1017 DDDMGDLVDAEYLYPQCGFFCPDPAAGAGMHHRRSSSRSGGDLTGLPSEBEA 1076  
 DB 1011 DDDMGDLVDAEYLYPQCGFFCPDPAAGAGTAAHRRSSSRSGGDLTGLMESPGE 1070  
 QY 1077 PPSPLAPSEAGSDVFDGDLGGAAGKGLQSLPTHDPSPLOKRYSEDPVPLPSETDGYVAP 1136  
 DB 1071 PPSPLAPSEAGSDVFEGLGKATKGPQGISPRDLSPLOKRYSEDPVPLPSETDGYVAP 1130  
 QY 1137 LTCSPPEYVNPQVPRPQPPREGLPLAPAPAGATLEBAKTLSSGKGVVMDVAFGGA 1196  
 DB 1131 LACSPPEYVNPQVPRPQPPREGLPLAPAPAGATLEBAKTLSSGKGVVMDVAFGGA 1190  
 QY 1197 VENPEYLTPOGGAPOPHPPAFSPAFDMLYWMQDPPERGAPESTFGKTPAENPEYLG 1256  
 DB 1191 VENPEYLTPOGGAPOPHPPAFSPAFDMLYWMQDPPERGAPESTFGKTPAENPEYLG 1249  
 QY 1257 LDVFPV 1261  
 DB 1250 LDVFPV 1254

## RESULT 4

## EGFR HUMAN

ID EGFR\_HUMAN STANDARD; PRT; 1210 AA.  
 AC P00533; P06268; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;  
 AC 000688; Q9B8S2; Q9B8C9; Q9GZK1; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor  
 protein-tyrosine kinase ErbB-1).  
 GN EGFR OR ERBB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID:9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=84219729; PubMed=6328312;  
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,  
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;  
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant  
 RT expression of the amplified gene in A431 epidermoid carcinoma cells";  
 RL Nature 309:418-425(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RA Tlekis J.V., Stark B.C., Scoccia B.;  
 RT "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta";  
 RL Mol. Reprod. Dev. 41:149-156(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RA Reiter J.L., Mainle N.J.;  
 RT "A 1.8 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor";  
 RL Nucleic Acids Res. 24:4050-4056(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RA MEDLINE=97256547; PubMed=9103388;  
 RA Tlekis J.V., Gariti J., Niederberger C., Scoccia B.;  
 RT "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGFR) in ovarian cancer";  
 RL Gynecol. Oncol. 65:36-41(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RX TISSUE=Placenta;  
 RL MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 RA Schehl Sinclair C., Pearall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,  
 RA Mainle N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGFR transcripts encoding truncated receptor  
 RT isoforms";  
 RL Genomics 71:1-20(2001).  
 RN [6]  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reiter J.L., Thredgill D.W., Danielson A.J., Schehl C.M.,  
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,  
 RA Mainle N.J.;  
 RT "Human and mouse alternative EGFR transcripts encoding only the  
 RT extracellular domain of the receptor";  
 RL Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
 RT "Expression cloning of human EGF receptor complementary DNA: Gene  
 RT amplification and three related messenger RNA products in A431  
 RT cells";  
 RL Science 224:843-848(1984).  
 RN [8]  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I., Sullivan M., Wilson R.K., Ma D.P.,  
 RT "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAs overproduced in A431 carcinoma cells";  
 RL Nature 309:806-810(1984).  
 RN [9]  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Simmen F.A., Gope M.L., Schulz T.Z., Wright D.A., Carpenter G.,  
 RA O'Malley B.W.;  
 RT "Isolation of an evolutionarily conserved epidermal growth factor  
 RT receptor cDNA from human A431 carcinoma cells";  
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).  
 RN [10]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
 RA Waterfield M.D.;  
 RT "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription";  
 RL Oncogene Res. 1:375-396(1987).  
 RN [11]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RT "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RT factor receptor proto-oncogene RNA synthesis";  
 RL J. Biol. Chem. 266:1746-1753(1991).  
 RN [12]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Strutton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RT "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN [13]  
 RP SEQUENCE OF 540.  
 RA Kohda D.;  
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RN [14]  
 RP RECEPTOR ACTIVITY.  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Mroczkowski B., Mosig G., Cohen S.;  
 RT "ATP-stimulated interaction between epidermal growth factor receptor

RT and supercoiled DNA.";  
 RL Nature 309:270-273(1984).  
 RN [15]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,  
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.  
 RT Identification of a novel site in EGF receptor.";  
 RL J. Biol. Chem. 264:10667-10671(1989).  
 RN [16]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-126; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RX MEDLINE=96399132; PubMed=8962717;  
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 RT the epidermal growth factor receptor expressed in Chinese hamster  
 RT ovary fibroblasts.";  
 RL Growth Factors 13:121-132(1996).  
 RN [17]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; PubMed=10731668;  
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor  
 RT receptor.";  
 RL J. Biochem. 127:65-72(2000).  
 RN [18]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=98225196; PubMed=9556602;  
 RA Abe Y., Otake M., Inagaki F., Lax I., Schlessinger J., Kohda D.;  
 RT "Disulfide bond structure of human epidermal growth factor receptor.";  
 RL J. Biol. Chem. 273:11150-11157(1998).  
 RN [19]  
 RP REVIEW.  
 RX MEDLINE=87297456; PubMed=3039909;  
 RA Carpenter G.;  
 RT "Receptors for epidermal growth factor and other polypeptide  
 RT mitogens.";  
 RL Annu. Rev. Biochem. 56:881-914(1987).  
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF  
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is  
 CC involved in the control of cell growth and differentiation.  
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 CC secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/  
 CC truncated isoform/TEGR, 3/p110 and 4; are produced by  
 CC alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 CC expressed in ovarian cancers.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL; X00588; CA25240.1; -  
 CC DR EMBL; U95089; AAB51063.1; -  
 CC EMBL; U48722; AAC50802.1; -

DR EMBL; U48723; AAC50804.1; -  
 DR EMBL; U48724; AAC50796.1; -  
 DR EMBL; U48725; AAC50797.1; -  
 DR EMBL; U48726; AAC50798.1; -

Query Match 45.4%; Score 3108; DB 1; Length 1210;  
 Best Local Similarity 49.6%; Pred. No. 46-159;  
 Matches 633; Conservative 169; Mismatches 351; Indels 124; Gaps 26;

11 LLLALLPQAA--STQVCTGTDMKRLRPAFPTLDMRLHYOGCGVGNLETLFPTN 68  
 14 LLAALCPARALEBKVCQGTNSKLTQLCTFEEDHLSLQRMNCEVNLGNEITVQGN 73  
 69 ASLSPLQDIQEVGVYLLAHQVROVPLQRLRIVEGTQLFEDNYALAVLDNGDPLNTP 128  
 74 YLSFLKTIQEVAGVYLLALNTVERIPLENLQIGNMVYENSALAVLSND----- 126  
 129 VVGASFGGLREQLSLTEILKGVLIQNPQLQVDTLMDIFPKXNQLALLTIDTR 188  
 127 ---ANKTGKLELPMRLQELHGAVFSSNNPALCVESIQWRDIVSSDLFNSMSDFQNR 183  
 189 SPACHPCSPMCKGRSCWSESDCSLTFTVAGGCA-RCKGPLPTDCCHECCAGCTGP 247  
 184 LSCCKQCDPSGNGSCWGAENQCKLTILCAQCCSGRCRKSPSDCCHNCAGCTGP 243  
 248 KASDCLACHFNHSGICEALCPALVTYNTDFESPNRPEGRTFGASCTTACPVYLTSD 307  
 244 RSDCLVCCKRFDEATCKDTCPLMLYNPTTYQMDVNBEGKVSFATCVKCKPRNVYTD 303  
 308 VGSCTLVCPLEHQEATDGTORCEKSPCARVCYGQMYIKANSKFIQTELE-FAG 366  
 304 HSCYRACADSYEM-EDGVKCKCKCEPCCKVNGGIGSFK-DLSLNTNTHKFN 361  
 367 CKRTFGSLAFPEPSDGPASNTAPLQPEQLQVFTLEITGYLYISAPDSLPLSVFQ 426  
 362 CTSISGDHLILPFAERGDSFTHTPLDQELDLKTVKEITGFLILQAMPENRTDLHARE 421  
 427 NQVIRGRLHNGAVALTLQIGISWLGRLSRLREGSLALIHNTLCLFVTVRWDLF 486  
 422 NLEIRGRTKQHGQSLAVSLNTISLGRSLKETSDDVITSGKNKCYANTIMKKLF 481  
 487 RNPQALLHTANRPDECEVGEGLACHQLCARHGCWGPPTQCVNCSQFLRQCEVECHV 546  
 482 GNSGQKTKIISRGNSCKATQVCHALCSPEGCMGPBRDVCGRNVSREGCEVDKML 541  
 547 LOGLPREYNARHCIPCHPECOPOGASVTCGPEPADQVCAAHYQDPPCVARCFP--- 602  
 542 LEGEPREFEVENSECTIQCHPECLPQAMNITCTGRGPNCTIQCAHYIDGPHCVKTCGAGWG 601  
 603 -NNFTVSPFLRVKVSASHLEPDEBGAOCPINCTHSCVLDKCGPAEQASPLSTIV 661  
 602 ENNTLV-W---KVADAGHV-----CHLCPKCTTCTGCTGEGLEGCTNGPKIP--SIA 647  
 662 SAVVG--ILLVVLGVVFGILLIKRQOKIRKTYRRLLOETELVEPLTPSGAMNQAM 718  
 648 TQMVGALLILLVALGIG--LFMRRIIVRKRTRLRLOEBELVEPLTPSGEAPNQLL 704  
 719 RLKXTELRKVVVGSAGFGTVYKGIWIPDENYKIPAIKVLRENTSPKANKETLDBAY 778  
 705 RLKTEFERKIVLSSGAFGVYKGLMIPBGKVKIPAIKELRATSPKANKETLDBAY 764  
 779 VMAVGSPYVSRLLGICLTSTVQVLTQMPYGCILDHVENRGRGLSGODLNMCMQIAKG 838  
 765 VVASVDNPHVCRLLGICTSTVQVLTQMPYGCILDYRHKNDNGISGQYLLNMCQIAKG 824  
 839 MSYLEDVAVLRDLAARVLYKSPNHVYKITPGLARLLDIDETRYHADGVPIPKWALE 898  
 825 MNYLEDRLVRDLAARVLYKTPQVYKITPGLAKLGAEBEKEVHAGGVPIPKWALE 884  
 899 SILRRFTHOSDWSYGVTVWELMTFGKAPYDGIIPAREIPMLLEGEERLPOPTCTIDVY 958  
 885 SILHRIYHOSDWSYGVTVWELMTFGSKPFDGIPASISITLKEGRLLPOPTCTIDVY 944



```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 538 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690520D46FD2D2F5 CRC64;

```

Query Match 45.0%; Score 3080; DB 1; Length 1210;  
 Best Local Similarity 49.5%; Pred. No. 1.3e-157;  
 Matches 635; Conservative 161; Mismatches 360; Indels 128; Gaps 27;

```

QY 11 LLLALLPFGAA-STQVCTGTDMKRLPASPETHLDMRLHYOGCQVQVGNLTLYLPTN 68
DB 14 LITLCAAGALBEKKYCOGTSNRLTQIGFEDHFLSQRYNNCEVLLGNLETTYQNRN 73
QY 69 ASLSFLDIOEVOGVVLIHNGQVAVPLQRLRYRGVLPEDNVALAVDNGDPLNNTTP 128
DB 74 YDSFLKTIQVAGVLIATLVTRIPLENQITIRGALVYNTALAILSN----- 124
QY 129 VTGSPGGLRELQRLSTLEIKGVLIQRNPOLCYODTILMKDI----FHKNOALATLI 184
DB 125 -YGNRTGLRELPMRNQELILGAVRFSNNPILCNMTIQMRDIYGVVFNISMMDL--- 180
QY 185 DTNRSRACHPSCPMCKSGSRGWSSEDOCSLTRVCGGCA-RCKGPLPTDCHEGCAAG 243
DB 181 -QSHPSSCPKCDPSCPGSCMGGEENCOKLTKLIQAQCSHRGRSPSDCCGNQCAAG 239
QY 244 CTGRHSDCLACTLFFNHSIGICELHCPALVTYNTTFSMPNPEGRYFGASCYACPYNY 303
DB 240 CTGRRESDDLVQCKQFQDEATCKDTCPEMLNPTTYQMDVNPBEKYSFGATCYKCEPRNY 239
QY 304 LSTVSGCTLVCPHANOVTAEADGTQCEKSKPCAVCYGLGQVYIKANSKRTIGITELE 363
DB 300 VVTDHSGCVACGDDYEV-EEODIRCKCKCDGCRKACGIGIGERK-DTLSTINATNIK 357
QY 364 -FAGCKKIFGSLAFPEPFGDDPASNTAPLOPELOVFEETLEETGLVYISAPDSLPDL 422
DB 358 HFKCTAISGDHLPLFAFKDSTTRTPPLDPRELEIKTVKELTGLLIQAMPDWTDL 417
QY 423 SVFQNTLOVIRGRILHNGAYSILLOGIGISWLGRLSELGSLALIHNTLGFVTPV 482
DB 418 HAFENLEIIRGRTKQHQFSLAVVGNITSLGRSLKEISDGVYIISGNNTLYANTINW 477
QY 483 DQLFNRHQAALHTANRPEDECVGEGLAGHQLCARHGCMRPGTQCNGQCFRGQCEV 542
DB 478 KKLFGTNPQTKTKIMNABEKDCAVNHVNCPLSSSEGCWPEPRDVCSCNVSREGREVE 537
QY 543 ECRVLOGLPREYVVARHCLPCHPECOFONGSVTCFPEADQCAVCAHYKDPFCVARCP 602

```

```

DB 538 KCVILREBPREFENSECICQCHPECPLQMANNTICTGRGPNQCQCAHYIDGRCVKTCPA 597
QY 603 -----NNFTVSFWLRVPKVASHLBPDEGACQPCPINCHSCVDDDKCCPAGASPL 657
DB 598 GINGENNTLV--WKYA-----DANNVCHLANCTYGCAGGGLQCEVWPSGPKI 645
QY 658 TSIVSAVAGILLVVLGVVFGI-LIKROOKIRKYMRLLOETLEVEPLTPSGAMPNQ 716
DB 646 PSIAITGIVGGLFIV-VALGIGLFMRHRHIVKTRRLQREIVEPLTPSGEAPNQ 704
QY 717 QMILKETELRKVUGSAGFGVYVGMIPDEENKIPATVLAENLSPKANKELDE 776
DB 705 HRLIKETEKTKIVGSGAFGVYGLMIPDEEKVIPAALKELEASPKANKELDE 764
QY 777 AAYMAGSPYVSRLLGICLSTVQVLTQMLPAGCLLDHRENGRLGSDLLMCMQIA 836
DB 765 AAYMAVDNHNHVRLLGICLSTVQVLTQMLPAGCLLDHRENGRLGSDLLMCMQIA 824
QY 837 KGMSTYEDVRLVHDLAARNVLYKSPNHYKITDFGLARLLIDETEVHADGKVPARKMA 896
DB 825 KGMNVLDEDRVLVHDLAARNVLYKTPQHYKITDFGLAKLGAEEKEYHAGCKVPARKMA 884
QY 897 LESILRRPFGSDVWSYGVYVWELTFGAKPYDGI PAEIPDLLEKGRLPQPICTID 956
DB 885 LESILRIYTHOSDWSYGVYVWELTFOSKPYDGI PASDISILEKGERLPQPICTID 944
QY 957 VYMIWYKCMWIDSECRPRERELVSFERNAPDQREFTVIO-NEDLGASPLDSTFYRSLL 1015
DB 945 VYMIWYKCMWIDSDSPKRELLEFSKMAADPQRLVIOGDERMHLPSPTDSNFRALM 1004
QY 1016 EDDMDGLDVADEEYLVPOQGFPCPDPAAGMWHHRHSSSTRSGGGLITGLPSEEB 1075
DB 1005 DEEDMDVDVADBYLLIPQGF-----NSPST----- 1031
QY 1076 APRSPAPSEAGSDVDFDDLGAKAGLQSLPTHPSPLORESDEPTVLPSEET--DGY 1133
DB 1032 -SRPLPLSLSATSN-----NSTYACINRNGSCRYVEDALQRYSSPTGAVEDNIDDAF 1086
QY 1134 VAPLTSPOPEYVNOVDVFPDPPSPREGPLPAARPGATLEBAKTLSPGKGVVXDVFAF 1193
DB 1087 L-----PYPEVYNQ-SVFKRPAGSYQNVYHNQPLHP-----APGRDLHYQN--PH 1129
QY 1194 GGAVENTYL-TPQGAAPQPPHPPAFSPAFNLYWMO-----DP-----PER 1236
DB 1130 SNAVGNPEYLNIAQ-----PTCLSSGFNSPALMIQKSHJWSLDNPDYQDDFPPE 1180
QY 1237 GAPPSYFKGPTAENPEYVGLDVP 1260
DB 1181 TKPNGIFKG-PTAENAEYLRVAP 1203

```

RESULT 6  
 ERB4\_HUMAN STANDARD; PRT; 1308 AA.  
 AC Q15303;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-cytosine kinase erbB-4 precursor (EC 2.7.1.112)  
 DE (P180erbB4) (tyrosine kinase-type cell surface receptor HER4).  
 GN ERB4 OR HER4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCB1\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=93189574; PubMed=8383326;  
 RA Plozman G.D., Coutouscou J.-M., Whitney G.S., Green J.M., Carlton G.W.,  
 RA Roy L., Neubauer M.G., Shoyab M.,  
 RT Ligand-specific activation of HER4/P180erbB4, a fourth member of the

RT epidermal growth factor receptor family";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=97476287; PubMed=9334263;  
 RA Elettus K., Corfas G., Paul S., Choi C.J., Rio C., Plowman G.D.,  
 Klagsbrun M.;  
 RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific  
 RT tissue distribution and differential processing in response to  
 RT phorbol ester.";  
 RL J. Biol. Chem. 272:26761-26768(1997).  
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NRG-  
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND  
 CC NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.  
 CC NOT ACTIVATED BY EGF, TGF- $\alpha$ , AND AMPHIREGULIN.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB  
 CC RECEPTORS (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. (shown here) AND JM-B;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER  
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED  
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND  
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN  
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,  
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,  
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS.  
 CC LUNG, SALIVARY GLAND, AND PANCREAS.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: L07868; AAB59446.1; --  
 DR HSSP: P11362; IFGK.  
 DR Genew: HGNC:3432; ERBB4.  
 DR MIM: 600543; --  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00069; kinase\_1.  
 DR Pfam: PF00757; Furin-like\_1.  
 DR Pfam: PF01030; Recept\_L\_domain; 2.  
 DR Pfam: PF02757; YLP\_2.  
 DR ProDom: PD000001; Euk\_kinase; 1.  
 DR SMART: SM00261; Fur\_4.  
 DR SMART: SM00219; TyKc; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.  
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.  
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KM Alternative splicing.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.  
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).  
 FT TRANSSEM 652 675 POTENTIAL.  
 FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 186 334 CYS-RICH.  
 FT DOMAIN 496 633 CYS-RICH.  
 FT DOMAIN 718 985 PROTEIN KINASE.

FT NP BIND 724 732 ATP (BY SIMILARITY).  
 FT BINDING 751 751 ATP (BY SIMILARITY).  
 FT ACT\_SITE 843 843 BY SIMILARITY.  
 FT DISULFID 189 197 BY SIMILARITY.  
 FT DISULFID 193 205 BY SIMILARITY.  
 FT DISULFID 213 221 BY SIMILARITY.  
 FT DISULFID 217 229 BY SIMILARITY.  
 FT DISULFID 230 238 BY SIMILARITY.  
 FT DISULFID 234 246 BY SIMILARITY.  
 FT DISULFID 249 258 BY SIMILARITY.  
 FT DISULFID 262 289 BY SIMILARITY.  
 FT DISULFID 293 304 BY SIMILARITY.  
 FT DISULFID 308 323 BY SIMILARITY.  
 FT DISULFID 326 330 BY SIMILARITY.  
 FT DISULFID 503 512 BY SIMILARITY.  
 FT DISULFID 507 520 BY SIMILARITY.  
 FT DISULFID 523 532 BY SIMILARITY.  
 FT DISULFID 536 552 BY SIMILARITY.  
 FT DISULFID 555 569 BY SIMILARITY.  
 FT DISULFID 559 577 BY SIMILARITY.  
 FT DISULFID 580 589 BY SIMILARITY.  
 FT DISULFID 593 614 BY SIMILARITY.  
 FT DISULFID 617 625 BY SIMILARITY.  
 FT DISULFID 621 633 BY SIMILARITY.  
 FT MOD\_RES 1162 1162 BY SIMILARITY.  
 FT MOD\_RES 1188 1188 BY SIMILARITY.  
 FT MOD\_RES 1258 1258 BY SIMILARITY.  
 FT MOD\_RES 1284 1284 BY SIMILARITY.  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 181 181 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 253 253 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 410 410 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 473 473 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 548 548 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC). (POTENTIAL).  
 FT VARSPLIT 626 648 NGPTSHCIYPMWGSHSTLPPQHA -> IGSSIEDICIGLMD  
 (IN ISOFORM JM-B).  
 SQ SEQUENCE 1308 AA; 146807 MW; 5BAAB80985D88761 CRC64;  
 Query Match 43.1%; Score 2945.5; DB 1; Length 1308;  
 Best Local Similarity 44.9%; Pred. No. 2.2e-150;  
 Matches 609; Conservative 181; Mismatch 378; Indels 189; Gaps 30;  
 QY 9 WGLLLALLPPGAA-----STGYCTGTDMLRLPASPETHLDMRLHLYOCQYVQGNLEITY 64  
 DB 8 WMTWSLLVAAGTVPDSOSVCACTENKLSLSLDEQYRLRKRYENCEVVMNLEITS 67  
 QY 65 LPTNASISFLDIOEVQGYVLIHNOYRQVPLRLRIVRGTLQFEDNYALAVLDNGDEPLN 124  
 DB 68 IEHNRDLSEFLRSVEVTVGYVLAINFOFYLPLEMLRIIRGKTYEDRYALAIPIFYNRKD 127  
 QY 125 NTTPTVGTASPGGLRELORSITLILKGVVLIQRNPOLCYOITLIMKDIFFHNGDLATLI 184  
 DB 128 NF-----GLQELGLANLTLEILNGCVYVDQNKFLCYADITIMQDVRNPPASNTLIV 178  
 QY 185 DTRSRFACHPCSPMKGRCSGESSSEDQSLTRVYAGGC-ARCKGPLPTDCHQCAAG 243  
 DB 179 STNGSGGCRCHKSCTG-RCWGPTEHNCQTLTRVYCAQCGRGVPVSDCHRECAAG 237  
 QY 244 CTGPRHSCCLACLNHNSGICELHCPALVYNTDTFESMPREGRITFGASCVYACPNY 303  
 DB 238 CSGPDTQCFACQMFNDGACVTCQPTQFVNPFTTFQLEHFNKAYTYGACVCKCPNF 297  
 QY 304 LSTDVSGCTLVCLPHNCEVTAEDGTORCEKSCPCARVCYGLAQVIKANSKFTIGITELE 363  
 DB 298 V-VDSSSCVRAQSPSSKKEV-BENGIKMKCPCTDICKAKCDGIGISGLMSAQTVSSNIDX 355  
 QY 364 FAGCKITGSLAFLEPESFDGDPASNTAPLQEPQOVETLEITIGLYIYSAMPDPLSL 423



```

Db      .356 FINCTKNGNLIIFVTGHDPYNAIEAIDEKLNIVRTVEITGFLNIGSWPMNMTDFS 415
Qy      424 VFCNLQVIRGRILHNGAVSLTLQGLISLWLGRLSRLRELSGLALIHNTHLCEFTVPMW 483
Db      416 VFSNLVITIGRVLVLSGLSLILLKQGGITSLOFQSLKELSAKNITITDNSNLCYHTIMWT 475
Qy      484 QLFENPHOALHTANRPEDCEVGGGLACHOLCAGHCGMGPGTCCVNCQSLRQCEVEE 543
Db      476 TLFSTINQRIYIRNRRKACNCTAGMVCNHLCSDDGCGPDDCLSCRFSSRGICIES 535
Qy      544 CRVLQGLREVYVNAHCLPCHPECCP-ONGSVTCFGEADCCVACAYKPPPCVACCP- 601
Db      536 CNLVDGEREFENSGICVBCDPCQCBKMEDSLTLTHGROPDNCICSHKQPCNCEKCPD 595
Qy      602 ---FNNFTVSFWLRVPKVSASHLEPEEGACOPCPINCTHSCVLDLDDKGC----- 648
Db      596 GLQGANSEIFKY-----ADPRE--CHPCENCTGCGNGPTSHCITYYPWTGHS 642
Qy      649 --PAQRASPLTISAVV--GILLVVLGVVFGILLIKRQKIRKYMRLLOTELYEP 705
Db      643 TLPOHAR-TPL--IAAGVIGLFLVLTAVGLTFAYVVRKSKIK-KKRLRLRL-ETELVER 697
Qy      706 LTPSGAMENQOMELKTELKRYKVLGSGAFGVYKGIWIPDENYKIPYALKVLRNT 765
Db      698 LTPSGTAPNQQLILKTELKRYKVLGSGAFGVYKGIWIPDEYKIPYALKVLRNT 757
Qy      766 SPXANKELDEAYMAGVSPYVSRLLGICLTSTVOLYQTMYPGCLLDHYRENGRLGS 825
Db      758 GPKANVEFMDALMASDHPLVRLGLVCLSPICVLTQMLPGCLLEVYHEKMDIGS 817
Qy      826 QDLNMCQIAKGSYLEDVRLVHDLAARNVLYKSPHNYITDPGLARLLDDETEYHA 885
Db      818 QLLNMCQIAKGSYLEDVRLVHDLAARNVLYKSPHNYITDPGLARLLDDETEYHA 877
Qy      886 DGVVPIKMALESILRRFTHSDVMSYGVTVLMTFGAKPYDIPAREIPDLLEKGE 945
Db      878 DGVVPIKMALESILRRFTHSDVMSYGVTVLMTFGAKPYDIPAREIPDLLEKGE 937
Qy      946 RLPOPICTIDVYIMVCMWIDSECRPRELIVSEPSRMAADQRFVIVONED-LGPAS 1004
Db      938 RLPOPICTIDVYIMVCMWIDSECRPRELIVSEPSRMAADQRFVIVONED-LGPAS 997
Qy      1005 PLDSTFYRSLLEDMDMDLYDAEEYLVPOQGFPCDPAPGAGVHHRRHSSSTRSGGD 1064
Db      998 PNDKFPQNLLEDMDMDLYDAEEYLVPOQGFPCDPAPGAGVHHRRHSSSTRSGGD 1046
Qy      1065 LTGLPSESEBAPPS-----PLAP-SEGAGSDVFDGD 1095
Db      1047 -EIGHSPPPAYTPMSGQFYRDGFAEGGVSVYRAPPTSTIPBAVAGGATAEITFDDS 1105
Qy      1096 LGMGAAGLOSLPFDSPSLQRYSEDEPTVPLPS-----ETDGYVAPLTCSPQPEYVQ 1148
Db      1106 CCGNGLRKVPYAPHQDESDSTQRYADPTVAPERSPRGELDEEYIMTPMDKXQKQELNP 1165
Qy      1149 PDVFPQPPSPREGGLPAPAPGATLBERAKTLSPCKNGVAVKDVAFGAVENPEYLTPOGG 1208
Db      1166 VE-----ENPFVSR-----KNGDLO-----ALNPEYHNASNG 1194
Qy      1209 AAPQHPPPA-----FSPAPNLVYVQDDP 1234
Db      1195 -----PPKADEYVNEPLYINTFANTLGAAYIKNNIISMPKAKAFBNPDYMHSLP 1248
Qy      1235 ERGA--PSTFFKGTPT-----AENDEYL 1255
Db      1249 PRSTLQHPDYLOEYSTKYFYKONGRIRPIVAENPEYL 1285

```

RESULT 7  
 ERB4\_RAT  
 ID ERB4\_RAT STANDARD; PRT; 1308 AA.  
 AC Q62956; Q922N7;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)

```

DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN      ERB4 OR TYRO-2.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
ON      NCBI_TaxId=10116;
RX      MEDLINE=98221155; PubMed=9553078;
RC      TISSUE=Heart;
RA      Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
RA      Marchionni M.A., Kelly R.A.;
RT      "Neuregulins promote survival and growth of cardiac myocytes.
RT      Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT      ventricular myocytes."
RL      J. Biol. Chem. 273:10261-10269(1998).
RN      [2]
RP      SEQUENCE OF 848-901 FROM N.A.
RC      TISSUE=Sciatic nerve;
RA      MEDLINE=91222560; PubMed=2025425;
RT      Lai C., Lemke G.;
RT      "An extended family of protein-tyrosine kinase genes differentially
RL      expressed in the vertebrate nervous system."
RN      Neuron 6:691-704(1991).
RP      SEQUENCE OF 1031-1198 FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX      MEDLINE=97184212; PubMed=9030624;
RA      Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT      "Expression of neuregulins and their putative receptors, Erb2 and
RT      Erb3, is induced during Wallerian degeneration."
RL      J. Neurosci. 17:1642-1659(1997).
CC      -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS. NRG-
CC      2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC      NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC      NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC      RECEPTORS (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC      HEART.
CC      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC      RESIDUES (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      This Swiss-Prot entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC      The European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
DR      EMBL; AF041838; AAD08899.1; -
DR      HSSP; U52531; AAC53051.1; -
DR      HSSP; P11362; 1FGK.
DR      InterPro; IPR000494; EGFR_L domain.
DR      InterPro; IPR000719; Euk_Dkinase.
DR      InterPro; IPR002174; Furin-like.
DR      InterPro; IPR001245; Tyr_pkinase.
DR      InterPro; IPR004019; YLP motif.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF00069; pkinase; 1.
DR      Pfam; PF01030; Recep_L-domain; 2.
DR      Pfam; PF02757; YLP_2; 1.
DR      PRINTS; PR00109; TYRKINASE.
DR      ProDom; PD000001; Euk_pkinase; 1.

```

DR SMART: SMO0261; FU: 4.  
 DR SMART: SMO0219; TYRC: 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
 FT SIGNAL 1 25 POTENTIAL.  
 FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.  
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 652 675 POTENTIAL.  
 FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 186 334 CYS-RICH.  
 FT DOMAIN 496 633 CYS-RICH.  
 FT DOMAIN 718 985 PROTEIN KINASE.  
 FT NP\_BIND 724 732 ATP (BY SIMILARITY).  
 FT BINDING 751 751 ATP (BY SIMILARITY).  
 FT ACT\_SITE 843 843 BY SIMILARITY.  
 FT DISULFID 189 197 BY SIMILARITY.  
 FT DISULFID 193 205 BY SIMILARITY.  
 FT DISULFID 213 221 BY SIMILARITY.  
 FT DISULFID 217 229 BY SIMILARITY.  
 FT DISULFID 230 238 BY SIMILARITY.  
 FT DISULFID 234 246 BY SIMILARITY.  
 FT DISULFID 249 258 BY SIMILARITY.  
 FT DISULFID 262 289 BY SIMILARITY.  
 FT DISULFID 293 304 BY SIMILARITY.  
 FT DISULFID 308 323 BY SIMILARITY.  
 FT DISULFID 326 330 BY SIMILARITY.  
 FT DISULFID 503 512 BY SIMILARITY.  
 FT DISULFID 507 520 BY SIMILARITY.  
 FT DISULFID 523 532 BY SIMILARITY.  
 FT DISULFID 536 552 BY SIMILARITY.  
 FT DISULFID 555 569 BY SIMILARITY.  
 FT DISULFID 559 577 BY SIMILARITY.  
 FT DISULFID 580 589 BY SIMILARITY.  
 FT DISULFID 593 614 BY SIMILARITY.  
 FT DISULFID 617 625 BY SIMILARITY.  
 FT DISULFID 621 633 BY SIMILARITY.  
 FT MOD\_RES 1162 1162 PHOSPHORYLATION (AUTO-).  
 FT MOD\_RES 1188 1188 PHOSPHORYLATION (AUTO-).  
 FT MOD\_RES 1258 1258 PHOSPHORYLATION (AUTO-).  
 FT MOD\_RES 1284 1284 PHOSPHORYLATION (AUTO-).  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC).  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC).  
 FT CARBOHYD 223 223 N-LINKED (GLCNAC).  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC).  
 FT CARBOHYD 410 410 N-LINKED (GLCNAC).  
 FT CARBOHYD 473 473 N-LINKED (GLCNAC).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC).  
 FT CARBOHYD 548 548 N-LINKED (GLCNAC).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC).  
 FT CONFLICT 1062 1062 S -> N (IN REF. 3).  
 FT CONFLICT 1080 1082 PVT -> SYR (IN REF. 3).  
 SQ SEQUENCE 1308 AA; 146957 MM; D944BB096A08B41 CRC64;

Query Match Best Local Similarity 42.8%; Score 2929; DB 1; Length 1308;  
 Matches 606; Conservative 188; Mismatches 388; Indels 176; Gaps 29;

QY 1 MELA-ALCRGILL--ALLPGAASVCTGTDMKLRIPASPEHLLDMLRLHYGCGCVQV 57  
 DB 1 MCLATGLWVGSLVAARTVQPSASQSCAGTENTLSLSDLEQYRALRKYEENCEVVM 60

QY 58 GNLELTLYPTNASLFLDIOEVGVYLIAHNOVQVLOLRIRYRGTOLEEDNYALAVL 117  
 DB 61 GNLELTLSLEHNDLFLRSIRKVTGIVLALNQFYLPDENRIRIGKLYEDRALALIF 120

QY 118 DNGDPLNNTTPTVLTGASPGGLRELQRLSLTEILKSGVLIQRNPOLCYQDTILMKDIFHGN 177  
 DB 121 LNYRKGNF-----GLQELGLKNTLEILNGGVVYDQNKFLCYADTIHMDIVRPFM 171

QY 178 QLALLTIDTNRSRACHPCSPMKSGSRGSESSDCCSLTRVCAGGC-ARCKGPLRTDCC 236  
 DB 172 PSMNTLVSTIGSSGGRCHKSCG-RCKGTEHNCQTLRTVACEDGDCRCYPRVSDCC 230

QY 237 HEQCAAGCTGPKHSDCLACHFNHSICELHCPALVTYNTDTESMPNREGRTFGASCV 236  
 DB 231 HRECAGGCGSPKTDTCFACMNFDSGACVTCQCFQTFVYNTTFQLHFNPAKTYGAFCV 230

QY 297 TACPYYLSTDVSCVLVGLAHQETAEADGTORCEKSGPCARVCYCGMGYIKANSKF 356  
 DB 291 KCPHNFV-VDSSCVRACPSCKMEV-EEHGIIMCKPCDTCIKACDQIGTGLMSAQTV 348

QY 357 IGITTEFAACCKIFGSLAFLPESFGDDPASNTAPQPEQVFEETLEETGLYISAMP 416  
 DB 349 DSNINIKFICTKINGNLLFLVTGIGHDPYNAIDAIDPEKLVNFRVREITGLNIQTWP 408

QY 417 DSLPDLSPVNPQVITGRILHNAYSLTLOGLSIWLGLSLRELSSGLALHNNHLCF 476  
 DB 409 PNMTDSVFNLTVIGRVLVSGLSLLIKQOITSLQFSLKEISAGNIYITDNSLCY 468

QY 477 VHTVPMQDLFRNPHQALLHTANRPEDECVGELACHQLCARGHCMGPGPTQCNCQFLR 536  
 DB 469 YHTINMTLTFSTYNOQIVIRDNRAENCTREGVACHLCSNDGCMGPGPQCSCRSFR 528

QY 537 GQECVEECRVLOGLPREYVNAHCLPCHPECQ-PQNGSVTCFGEADQVCAHYKDPF 595  
 DB 529 GKICIESCNLYDEDEFEFENGSGICVECDSCCEKMEBGLTLCHGPGPDNCTCKSHFXDGN 588

QY 596 CVARCP-----FNNFTVSWFLRVPKVASLSLPEDEGAQCPICNTHSCVDDIDGCG-- 648  
 DB 589 CVERCPDVLOGANSIFKY-----ADQRECHPCHPNCTQCGNGTSHDCTY 635

QY 649 -----PAQRASPLTISYSAVY-GILLVVLGVFGLIKRQOKIRKYMRLQ 698  
 DB 636 YPMTGSHTLPHQAR-TPL--IAGVGVLGVFLVTALTFVAVYRRKSIX-KKRALRRFL- 690

QY 699 ETELVEELTSSGAMPQAOQRIKTELRLKRVKLSGAGFTYVKGIVWPGETVKIPAI 750  
 DB 691 ETELVEELTSSGAMPQAOQRIKTELRLKRVKLSGAGFTYVKGIVWPGETVKIPAI 750

QY 759 KYLRENTSPLANKEILDEAVVMAGSPVYSRLGICLSTVOLVQLMAYGCLLVHRE 818  
 DB 751 KILNETTPRANVEFDEALIMASVHPHVRLVGLCSTTIOLVQLMHGCLLEVHE 810

QY 819 NRGRLSGQDLLNRCMOJAKMSYLEDVRLVHRDLAARNVLYKSPNHVKTDFGLARLDI 878  
 DB 811 HKDNIGQLLLNMCVQIAKGMVLEERLVRDLAARNVLYKSPNHVKTDFGLARLDI 870

QY 879 DETEYHADGKVEIKMMALESIARRFTHOSDVMSGVYWEIMTGAKPYDGIIPAREIP 938  
 DB 871 DEKEYNADGKMEIKMMALLECTHYRKFTHOSDVMSGVYWEIMTGAKPYDGIIPAREIP 930

QY 939 DLEKGRRLPQPICTIDVYIMVKCMMDISCRPPRELVSFSSMARDPQRFVYIQNE 998  
 DB 931 DLEKGRRLPQPICTIDVYIMVKCMMDISCRPPRELVSFSSMARDPQRFVYIQNE 990

QY 999 D-LGPAASPLDSTFYRLSLEDMDGDLVDAEYVLPQGFPCFDP----- 1041  
 DB 991 DRMLKLPENDSKFFQNLDEEDLDDMDDEEYVLP-QAEFTIPPIYTSRTIDSNSREIG 1049

QY 1042 -----APGAGGVHHRHRSSTRSGGDLTLGLEPSEEARSPPLASSEGADVDFG 1094  
 DB 1050 HSPRPAYTPMGSQFYVQDGFATQCG---MPPRYATNTSTIEPAVVA--QGATAMFDD 1104

QY 1095 DLGMAKAGKQSLPTHDPSFLQVYSEDPTPLPS-----ETDGVVAPLTGSPQPEYVN 1147  
 DB 1105 SCNGNGLRKRVVPHVQEDSSQRYSDPVTFAERNPRALDEGGYTPMHDKPKQKEYLN 1164

QY 1148 QPVRPQPPSPRGRPLPAARPAAGATLEAKTSLPGKNGVYGVDFAFGAVENREYLTPOG 1207  
 DB 1165 PVE-----ENPVSRR-----KNGDLQ-----ALDNEFYHSAS 1193

QY 1208 GAAPQPHPPA-----FSPAFDNLVYWDQDP 1233

Db 1194 G-----PPKAEDEYNEPPLTFTNALGNAEYKNSLLSVEKAKAFDNDPDYMHSL 1247  
 QY 1234 PERGA--PSTFKGTFT-----AENPEYL 1255  
 Db 1248 PPSTSLHPDYLOEYSIKYFYKONGRIPIVAENPEYL 1285

RESULT 8  
 XMRK\_XIPMA STANDARD; PRT; 1167 AA.  
 ID XMRK\_XIPMA  
 AC P1338;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).  
 GN XMRK OR TU.  
 OS Xiphophorus maculatus (Southern platyfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Perciformes; Atherinomorpha;  
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
 CX NCBI\_TaxID=8083;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90015140; PubMed=2797166;  
 RA Meltschke J., Adam D., Malitschek B., Maueker W., Raulf F.,  
 RA Telling A., Robertson S.M., Schartl M.,  
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-  
 RT inducing Tu locus in Xiphophorus.";  
 RL Nature 341:415-421(1989).  
 RN [2]  
 RP REVISION TO 515.  
 RA Schartl M.,  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBU databases.  
 CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.  
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein-  
 CC tyrosine phosphate.  
 CC - SUBCELLULAR LOCATION: Type I membrane protein.  
 CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.  
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; X16891; CA34770.2; -  
 DR PIR; S06142; S06142.  
 DR HSSP; P13362; 1FCG.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF000659; pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SMC0261; FU; 5.  
 DR SMART; SMC0220; S\_TKc; 1.  
 DR SMART; SMC0219; TYKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;  
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

FT DOMAIN 26 642 KINASE.  
 FT TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 666 1167 POTENTIAL.  
 FT DOMAIN 710 977 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 716 724 PROTEIN KINASE  
 FT BINDING 743 743 ATP (BY SIMILARITY).  
 FT ACT\_SITE 835 835 ATP (BY SIMILARITY).  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 220 228 BY SIMILARITY.  
 FT DISULFID 224 236 BY SIMILARITY.  
 FT DISULFID 237 245 BY SIMILARITY.  
 FT DISULFID 241 253 BY SIMILARITY.  
 FT DISULFID 256 265 BY SIMILARITY.  
 FT DISULFID 269 296 BY SIMILARITY.  
 FT DISULFID 300 311 BY SIMILARITY.  
 FT DISULFID 315 330 BY SIMILARITY.  
 FT DISULFID 333 337 BY SIMILARITY.  
 FT DISULFID 504 513 BY SIMILARITY.  
 FT DISULFID 508 521 BY SIMILARITY.  
 FT DISULFID 524 533 BY SIMILARITY.  
 FT DISULFID 537 553 BY SIMILARITY.  
 FT DISULFID 556 569 BY SIMILARITY.  
 FT DISULFID 560 577 BY SIMILARITY.  
 FT DISULFID 593 615 BY SIMILARITY.  
 FT DISULFID 618 626 BY SIMILARITY.  
 FT DISULFID 622 634 BY SIMILARITY.  
 FT DISULFID 114 114 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 144 144 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 201 201 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 356 356 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 365 365 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 398 398 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 621 621 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 38.6%; Score 2643.5; DB 1; Length 1167;  
 Best Local Similarity 44.6%; Pred. No. 2.8e-134;  
 Matches 569; Conservative 165; Mismatches 392; Indels 151; Gaps 30;

QY 4 AALCWGILLALPPGAASST---QVCTGTDMKRLPASPETHLMDLRLHYOGCOVQGN 59  
 Db 8 AALLQ--LLLVLSISRCSTDPDRKVCQGTSTNQM---LDVHYLKMKKMYSGCNVVLEN 62  
 QY 60 LETYLPFNASLSFQDIOEVGYVLIHNOVAVPQLRLRVRGTOLPEDNYALAVLDN 119  
 Db 63 LEITTOENODLSFIQSIQEVGYVLIHNEVSTIPLVNLRLIRQNLIEGNFTLLVMSN 122  
 QY 120 GDLNNTTPVTGASFGRLRELQRLSLEILKGVLIQNPOLCYODTILMKDIFKKNQL 179  
 Db 123 YQK-NPSSP--DVYVGLKQQLSLVLTLLISGVYVSHNPILCNVETINMMDIVDKTSNP 179  
 QY 180 ALLTLDTRSPRCHGCSMKGSRGWSSSSDQSLTFTVCAAGC-AACKGPLPDDCHE 238  
 Db 180 TMLLIPHAERQCQCDHGCYVNGSCWABGPHQCQFTYLLCAEQGNRCRGKPIDCCNE 239  
 QY 239 QCAAGCTGPKHSDCLACHFNHSGICELCPALVYNTDTFESMPNPGRTTFGASCYTA 298  
 Db 240 HCAGGCTGPBRATDCLAGDFNDGCTCKTCTPPKYLIVSHQVNDVNPXIKTFGACAYKE 299  
 QY 299 CPYNTLSDVGSCTLVCLHNOEYTAEGTQRCCKSPCARVCYGLGMOYIKANSKIG 358  
 Db 300 CPSNVVVE-GACVASCAGMLEVD-ENGKSCRCPCDVCPCVCGIGSL-SVTIVN 356  
 QY 359 ITEL-EFAGCKKIFISFLPESFDGPDASNTAPLPQPOLQVFTLEBITGLVYSAMPD 417  
 Db 357 STNIRSFNCIKINDIILNNSPBGDHYHYKIGMDEPHMLNLTIVKLTGLVLYMMPD 416  
 QY 418 SLPLSVFQNLQVIRGILHNGAVS-LTLQGLISWGLRSLRELGSGLALIHNTHLCF 476

```

Db 417 NMISLSVFNLEIRGTTTSRSGSFVVOYRHLQWGLASLSEVSGNVILKNTDQARY 476
Qy 477 VHTVPMQDLFRNPHQALLHTANRDECEVGBGLACHQLARGHCWGFPPQCVNCSOFLR 536
Db 477 ANTIMMRRLFRSEQSIEDYART-----ENQCNNECEGDEGCGWGPMTVCVSLHVD 529
Qy 537 GQEVCECRVYQGLPREYVNAARHCLPCHPECOPOGNSVTFEGPEADCCVACAHYKDPFC 596
Db 530 GGRCVASCNLQGFPRRQVDGRCVQCHQECVOTDSTLTGCPANCSAHFQDQPOC 589
Qy 597 VARCP---FNNFTVSFWLRVPKVSASHLEPDEBACQPCPINTHSCVDDDDKCPAQR 653
Db 590 IPRCPHGLDGDTLNWKYA-----DKMQCCQPCQNCQTCGSGGSLGCRD-1 638
Qy 654 ASPLTSYSAVGLLVVGVVGLILKRCQCKIRKYMRLQLGELYEPLTPSGAMP 713
Db 639 VSHSLVAVGLVGLITVALLIVLLRRRIR-KRRITRCLQEXELVEPLTPSQAP 697
Qy 714 NQAGRLKETELKRVKVLGSGAFGTYKGIIPDGENVAKPVALKYLENTSPKANKI 773
Db 698 NQAFRLIKETEFKKDLRSLGSGAGTYKGIWNGDNINIRPAIKVLRATSPKXQEV 757
Qy 774 LDEAYVNAVGSPPVSRLLGICLTSTVQLVQLMPPGCLLDHRENGRIGSQDLLNWC 833
Db 758 LDEAYVNAVDHPVCRLLGICLSAVQLVQLMPYGLLDYVQHGERICQGLNWC 817
Qy 834 QIAGKSYLEDVRLVHRDLAARNLVKSPNHKITTDLGLRLDIDETEVHADGKVPK 893
Db 818 QIAGKMYLEERHLVHRDLAARNLVKSPNHKITTDLGLRLDIDETEVHADGKVPK 877
Qy 894 WMALESILRRRTHQSDVSYGVTVWELMTFGAKPYDIPAREIPDLEKGERLPQPIC 953
Db 878 WMALESILQMTYTHQSDVSYGVTVWELMTFGSKPYDIPAREIPDLEKGERLPQPIC 937
Qy 954 TIDVYMTVMKWMIDSECRPRRELVSFSHMARDPQRFVYIQMEDIGRPSPLDSTYRS 1013
Db 938 TIEVYMTILKCMWIDPSRPRFRELVEGFSQWADPSRYLTQ---NLPSLSRRILFSR 994
Qy 1014 LLEDMDMDLVDAEYVLPQCGFPCPPAPAGAGVHRRHSSSTRSGGDLTLGLEPSE 1073
Db 995 LLSDD--DVYDADBEYLLPYRI-----NRQGS----- 1020
Qy 1074 EEARSPPLASEGSGSVDFDGLMGAKGLQSLPTDPSRLQYSDPY-PLPSTDG 1132
Db 1021 ---EPCIPPTG-----PVRENSTLTINISDPQNALEKDLG 1055
Qy 1133 YVAPLTCSPQPEYVNOVPVRPQ-----PSPRE-----GPLP-AARPAQATLERAKTL 1179
Db 1056 H-----EYVNPQSESTSRSLSDIYNVNYEDLTDGMPVSLSSQEAETNPSREYL 1105
Qy 1180 SPGXGVVGVVFAFGAVENPEYLTPOGGAAPQHPHPPAFSFAFDNTLYWDQDPEFGAP 1239
Db 1106 NTNONS--PLVSSGSMDDPY--QAG-----YQAF-----LPQAL 1140
Qy 1240 PSTFGKPTAEPEYL 1256
Db 1141 TGNMGFLPAENLEYLG 1157

```

```

RESULT 9
ERR3_HUMAN STANDARD; PRT; 1342 AA.
ID ERB3_HUMAN
AC P21860;
CD 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors."
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Ploman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene."
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=91282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase."
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT. HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M29366; AAA55790.1; -
DR EMBL; M34309; AAA55979.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; 1FGK.
DR Genew; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TyKc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.
DR PROSITE; PS00109; PROTEIN KINASE TYR; FALSE NEG.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; MultiGene family; Receptor; Signal;
DR Transferase; Tyrosine-protein kinase; Atp-binding; Phosphorylation;
KM

```

Alternative splicing.

Query Match	Score	DB 1	Length	Score	DB 2	Length
34.98	2387.5	DB 1	1342	34.98	2387.5	DB 2
Best Local Similarity	39.94	Pred. No. 1.7e-120		Best Local Similarity	39.94	Pred. No. 1.7e-120
Matches 527; Conservative 193; Mismatches 459; Indels 141; Gaps 34;				Matches 527; Conservative 193; Mismatches 459; Indels 141; Gaps 34;		

Query Match 34.98; Score 2387.5; DB 1; Length 1342;  
 Best Local Similarity 39.94; Pred. No. 1.7e-120;  
 Matches 527; Conservative 193; Mismatches 459; Indels 141; Gaps 34;

Query 10 GLTLLALPPGAA--STOVCTGTDMKLLPASPETHMLNHLVGGCQVQNGNLETLPT 67  
 Db 11 GLTLLALPPGAA--STOVCTGTDMKLLPASPETHMLNHLVGGCQVQNGNLETLPT 67  
 Query 68 NASISFLQDIQEVGVYLAHNOYQVPLQRLVRSQTQLEFEDYALAVLNDGPIUNTT 127  
 Db 71 NASISFLQDIQEVGVYLAHNOYQVPLQRLVRSQTQLEFEDYALAVLNDGPIUNTT 127  
 Query 128 PYVGASPGRLRELQRLTLLKGVLIQNPOLCQDITLMDIFHKKNQALLTIDIN 187  
 Db 126 ---NSSHALRQLTLLTLLSGVYIEKNDKICMHTDMDIYARD---AETVVK 178  
 Query 188 RSRACHPSPMCKSCRCMGSSSEDCSLTRTVGAGC-ARCKGFLPDCCHQCAAGCTG 246  
 Db 179 NGRSCPCPCHEVCKG-RCMGPGSEDCCLTCTICAPQNGHCPCGPNPQCHDCAGCGG 237  
 Query 247 PKHSDDLACLHPHNSGICELHCPALVTYNTDTESMNPBGRYTFGASCTACPYVLT 306  
 Db 238 PQTDFCACHFNDSGACVPRCPPLVYNKLTFLQLEPNPHTKTYQGVGVASCAPHNFV-V 296

Query 307 DVSGCTLVCPHNGQVTAEDGTORCEKSPKARVCYGLQMOYIKANSKC--IGITBLE- 363  
 Db 297 DVSGCTLVCPHNGQVTAEDGTORCEKSPKARVCYGLQMOYIKANSKC--IGITBLE- 363  
 Query 364 PACCKTFGSLAPLSEFQDDPASPANRPLQPELOVPELTETGYLYISANQSLPDL 423  
 Db 351 PACCKTFGSLAPLSEFQDDPASPANRPLQPELOVPELTETGYLYISANQSLPDL 423  
 Query 424 VFNQGLYIRGRIIANGAYS--LTQGLIGIWMGLSLRELQSGALLIHNHNLGFVHPV 482  
 Db 411 VFNQGLYIRGRIIANGAYS--LTQGLIGIWMGLSLRELQSGALLIHNHNLGFVHPV 482  
 Query 483 DQLEPRDQALLHTA-NRPEDECVGSLACHQLCARHCMPGPTQCVNCSQPLRGQEV 541  
 Db 471 DQLEPRDQALLHTA-NRPEDECVGSLACHQLCARHCMPGPTQCVNCSQPLRGQEV 541  
 Query 542 EECRVQLGPREVYNAHCLPCPECOQNGSVTCGPRADQVCAHYDPPFCVAPCP 601  
 Db 531 EECRVQLGPREVYNAHCLPCPECOQNGSVTCGPRADQVCAHYDPPFCVAPCP 601  
 Query 602 FNNFTVSPWLVRKVASHLPEDEGACOPCPINCHTSCVDLDDKGCAPQRA---SPL 657  
 Db 591 FNNFTVSPWLVRKVASHLPEDEGACOPCPINCHTSCVDLDDKGCAPQRA---SPL 657  
 Query 658 TSIVSAVAGLILVYVGVFGILIKRQCKIR-KYMRRLQETELVEPTBPGAMPNOA 716  
 Db 643 TSIVSAVAGLILVYVGVFGILIKRQCKIR-KYMRRLQETELVEPTBPGAMPNOA 716  
 Query 717 QMRILKETELRKVYLGSGAFGVYKIMPDGENYKIPALVYLRBENTSPANKETLDE 776  
 Db 700 QMRILKETELRKVYLGSGAFGVYKIMPDGENYKIPALVYLRBENTSPANKETLDE 776  
 Query 777 AYVMAVGSPYRSLIGICTSTVQVLTQMLPYGCLLDHYRENRGLSGQDLNMCQIA 836  
 Db 760 AYVMAVGSPYRSLIGICTSTVQVLTQMLPYGCLLDHYRENRGLSGQDLNMCQIA 836  
 Query 837 KGMVYEDVRLVYRDLAARVLYKSPNHYKIDPGLRLDIDETVHADGGVPIKMA 896  
 Db 820 KGMVYEDVRLVYRDLAARVLYKSPNHYKIDPGLRLDIDETVHADGGVPIKMA 896  
 Query 897 LESILRRFTQSDVNSYGVYVWELMTFGAKYDGIIPAREIDLEKGRLPQPCITD 956  
 Db 880 LESILRRFTQSDVNSYGVYVWELMTFGAKYDGIIPAREIDLEKGRLPQPCITD 956  
 Query 957 VYVIMYKCMWIDSECPREPRVSEPSRAAPORVYVQNEDLGPA---SPUDSTFYRS 1013  
 Db 940 VYVIMYKCMWIDSECPREPRVSEPSRAAPORVYVQNEDLGPA---SPUDSTFYRS 1013  
 Query 1014 LLEDDMDGLVDAAEYLVPOQGFPCDPAPAGAGMWHNRSSSTSGGGLTLGLEP-S 1072  
 Db 999 LLEDDMDGLVDAAEYLVPOQGFPCDPAPAGAGMWHNRSSSTSGGGLTLGLEP-S 1072  
 Query 1073 EEDAPRSLAPSEGAGDVFDGLGNAKGLQSLPTH-PSPLORYSEDPVPLP--- 1127  
 Db 1040 EEDAPRSLAPSEGAGDVFDGLGNAKGLQSLPTH-PSPLORYSEDPVPLP--- 1127  
 Query 1128 SETDQGYVA-----PLTSGQPE-----YVQNPVRPQPPRRGSP--- 1162  
 Db 1093 SETDQGYVA-----PLTSGQPE-----YVQNPVRPQPPRRGSP--- 1162  
 Query 1163 -----LPAAPGATLERAKTLP-GKNGV-----KDVFAFGAVENPEVLTPOG 1207  
 Db 1153 -----LPAAPGATLERAKTLP-GKNGV-----KDVFAFGAVENPEVLTPOG 1207  
 Query 1208 GAAPQHPHPPAPSPAFDNLVYD-----QDPPERCAPSTKGTFTANPEVL 1255  
 Db 1205 GAAPQHPHPPAPSPAFDNLVYD-----QDPPERCAPSTKGTFTANPEVL 1255

RESULT 10  
 ERB3 RAT  
 ID ERB3 RAT  
 AC Q62799; Q62955; STANDARD; PRT; 1339 AA.

DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)  
 DE (c-erbB3).  
 GN ERBB3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NC NCBI\_Taxid:10116;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=96096535; PubMed=8522190;  
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sterke S.L., Koland J.G.;  
 RT "Cloning of the rat ErbB3 cDNA and characterization of the  
 RT recombinant protein.";  
 RL Gene 165:279-284(1995).  
 RL (2)  
 RP REVISIONS TO 85, 513 AND 565.  
 RA Hellyer N.J., Koland J.G.;  
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 922-1097 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
 RX MEDLINE=97184212; PubMed=9030624;  
 RA Carroll S.L., Miller M.L., Frohman P.W., Kim S.S., Corbett J.A.;  
 RT "Expression of neurofilins and their putative receptors, ErbB2 and  
 RT ErbB3, is induced during Wallerian degeneration.";  
 RL J. Neurosci. 17:1642-1659(1997).  
 CC - FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTRK.  
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC - SUBCELLULAR LOCATION: Type I membrane protein.  
 CC - DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE.  
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 CC EMBL: U29339; AAC8498.2; -  
 CC EMBL: U52310; AAC3050.1; -  
 CC HSSP: P11362; 1FCR.  
 DR InterPro: IPR000494; EGFR\_L domain.  
 DR InterPro: IPR002179; Euk\_pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L domain; 2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00261; FU; 5-  
 DR SMART: SM00219; TYRK; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; FALSE\_NEG.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1339 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.  
 FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).  
 FT

FT TRANSMEM 644 662 POTENTIAL.  
 FT DOMAIN 663 1339 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 183 259 CYS-RICH.  
 FT DOMAIN 707 964 PROTEIN KINASE.  
 FT NP\_BIND 713 721 ATP (BY SIMILARITY).  
 FT BINDING 740 740 ATP (BY SIMILARITY).  
 FT ACT\_SITE 832 832 BY SIMILARITY.  
 FT DISULFID 186 194 BY SIMILARITY.  
 FT DISULFID 190 202 BY SIMILARITY.  
 FT DISULFID 210 216 BY SIMILARITY.  
 FT DISULFID 214 226 BY SIMILARITY.  
 FT DISULFID 227 235 BY SIMILARITY.  
 FT DISULFID 231 243 BY SIMILARITY.  
 FT DISULFID 246 255 BY SIMILARITY.  
 FT DISULFID 259 286 BY SIMILARITY.  
 FT DISULFID 290 301 BY SIMILARITY.  
 FT DISULFID 305 320 BY SIMILARITY.  
 FT DISULFID 323 327 BY SIMILARITY.  
 FT DISULFID 500 509 BY SIMILARITY.  
 FT DISULFID 504 517 BY SIMILARITY.  
 FT DISULFID 520 529 BY SIMILARITY.  
 FT DISULFID 533 549 BY SIMILARITY.  
 FT DISULFID 556 573 BY SIMILARITY.  
 FT DISULFID 576 585 BY SIMILARITY.  
 FT DISULFID 589 610 BY SIMILARITY.  
 FT DISULFID 613 621 BY SIMILARITY.  
 FT DISULFID 617 629 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 250 250 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 353 353 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 408 408 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 414 414 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 437 437 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 469 469 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 522 522 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 566 566 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 616 616 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).  
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F402BBDF5LE CRC64;  
 Query Match 33.9%; Score 2316.5; DB 1; Length 1339;  
 Best Local Similarity 40.4%; Pred. No. 1,1e-116;  
 Matches 521; Conservative 166; Mismatches 437; Indels 165; Gaps 35;  
 QY 3 LALCRWGLLLALLPQGA---STVCTGTDMLRIPASEFTLDMLRHYGCGVYQGN 59  
 DB 7 LQVLC---FLSLARSEWNGQAVCPGLNLSTVGADNQYQLYKYEKCEVWNGN 62  
 QY 60 LELTYPTNALSFLQIOEVGVYLAHQVQVPLQRIYRGTOLEFDNYALAVLDN 119  
 DB 63 LEIVLGHNDLSFLQIRIVETGVYLVANMFESVLPNLRVYRGTOVDGKRAIVM-- 120  
 QY 120 GDPINNTPTVYTGSPGGLRELRSLTEILKGVLLIQRPOLCYQDTILMKDIFKKNOL 179  
 DB 121 ---LNVNT---NSSHALRQLKFTQLTEILSGGVYIEKDKLCHEMTIDIRDIVRVR-- 170  
 QY 180 ALTLIDNRBRACHPSPMCKSGRCMGESSEDCSLTRTYCAGGC-ARCGPLPTCCHE 238  
 DB 171 GAETVKNNGANCPCHVEVCKG-RMGPEGDDQILIKTICAGQCNCRCGPFPNQCCHD 229  
 QY 239 QCAAGCTGPRHSDCLACLNHNSGICELNCPALVTYNTDFFESMPPEGRYTPGASCYA 298  
 DB 230 ECAAGGCGGPDTCFCACRPNDSGACVPCRPPELVYNNKLTFLQEPHPHTYGVGVAS 289  
 QY 299 CPNYVLTVDGSCVLYCPPLNNOVLTAEHQREKSKPARVCYL--GMQYIKANSKF 356  
 DB 290 CPNHFV-VDTFTVPCACPPDKMEVD-KHGKMCPEGGGLCPKXCEBTGSGSRQYTVSSN 347  
 QY 357 IGTLELFAACCKIFGSLAVLPSPFGDPAISNTAPLOPEQLQVFETLEITGLYISAMP 416  
 DB 348 ID----GFVYCTKILGNLDFLITGLAVNDPWHKIPALDPEKLVNFRIVREITGLYINQSNW 403  
 QY 417 DSLPDSVFPQNLQVITKGRILHNGAYS-LTLQGLIGISWLGRLSRLDELGSLALIHNTHLIC 475





RA Williams S.M., Woodage T., Morley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveri U.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gbbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*.";  
 RL Science 287:2185-2195(2000).  
 RN [7].  
 RP SEQUENCE OF 959-1078 FROM N.A.  
 RC STRAIN=Daekanyeong;  
 RX MEDLINE=85137938; PubMed=2983232;  
 RA Wadsworth S.C., Vincent W.S. III, Bilodeau-Wentworth D.;  
 RT "A *Drosophila* genomic sequence with homology to human epidermal  
 RT growth factor receptor.";  
 RL Nature 314:178-180(1995).  
 RN [8].  
 RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=92038942; PubMed=1936959;  
 RA Raz E., Schejter E.D., Shilo B.Z.;  
 RT "Interallelic complementation among DER/fib alleles: implications for  
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";  
 RL Genetics 129:191-201(1991).  
 RN [9].  
 RP REVIEW.  
 RX MEDLINE=97248481; PubMed=9094709;  
 RA Perrimon N., Perkins L.A.;  
 RT "There must be 50 ways to rule the signal: the case of the *Drosophila*  
 RT EGF receptor.";  
 RL Cell 89:13-16(1997).  
 CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,  
 CC WHICH IS AN ANTRAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAP-  
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.  
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE  
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL  
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE  
 CC ESTABLISHMENT OF CENTRAL CELL FATES. MAINTENANCE OF AMIOGEBROSA  
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE  
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF  
 CC CUTICLE.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I  
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE  
 CC PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; TYPE I (SHOWN HERE), TYPE II AND  
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: UBQUITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,  
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF  
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST  
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW. LEVELS REMAIN HIGH  
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN  
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX  
 CC AND THORACIC AND ABDOMINAL GANGLIA.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF052754; AAC08536.1; -;  
 DR EMBL; AF052753; AAC08536.1; JOINED.  
 DR EMBL; AF052754; AAC08535.1; -;  
 DR EMBL; AF052752; AAC08535.1; JOINED.  
 DR EMBL; K03054; AAA51462.1; -;  
 DR EMBL; K03417; AAA51460.1; -;  
 DR EMBL; K03416; AAA50965.1; -;  
 DR EMBL; K03418; AAA51461.1; -;  
 DR EMBL; AF109077; AAD26134.1; -;  
 DR EMBL; AF109078; AAD26132.1; -;

DR EMBL; AF109082; AAD26133.1; JOINED.  
 DR EMBL; AF109078; AAD26133.1; -;  
 DR EMBL; AF109084; AAD26133.1; JOINED.  
 DR EMBL; AF109079; AAD26130.1; -;  
 DR EMBL; AF109081; AAD26130.1; JOINED.  
 DR EMBL; AF109079; AAD26131.1; -;  
 DR EMBL; AF109083; AAD26131.1; JOINED.  
 DR EMBL; AF109080; AAD26135.1; -;  
 DR EMBL; AE003454; AAF46732.1; -;  
 DR EMBL; X02293; CAAG26157.1; -;  
 DR EMBL; X78920; CAAG55523.1; -;  
 DR EMBL; X78918; CAAG55521.1; -;  
 DR EMBL; X78919; CAAG55522.1; -;  
 DR PIR; A00640; GOFPE.  
 DR HSSP; P11352; 1PCK.  
 DR Flybase; FBgn0003731; Egfr.  
 DR Interpro; IPR000494; EGF\_L\_domain.  
 DR Interpro; IPR000719; Euk\_Pkinase.  
 DR Interpro; IPR002174; Furin-like.  
 DR Interpro; IPR001245; Tyr\_Pkinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_Ldomain; 2.  
 DR PRINTS; PR001001; TYRKINASE.  
 DR ProDom; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SM00261; FU; 7.  
 DR SMART; SM00219; TYRKC; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE\_TYR; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;  
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;  
 KW Developmental protein.  
 FT SIGNAL 1 30  
 FT CHAIN 31 1426 POTENTIAL.  
 FT DOMAIN 31 868 EPIDERMAL GROWTH FACTOR RECEPTOR.  
 FT TRANSMEM 869 889 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 890 1426 POTENTIAL.  
 FT DOMAIN 938 1198 CYTOPLASMIC (POTENTIAL).  
 FT NP BIND 944 952 PROTEIN KINASE.  
 FT BINDING 971 971 ATP (BY SIMILARITY).  
 FT ACT SITE 1063 1063 ATP (BY SIMILARITY).  
 FT MOD\_RES 902 902 BY SIMILARITY.  
 FT PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 Query Match 28.7%; Score 1960; DB 1; length 1426;  
 Best Local Similarity 32.5%; Pred. No. 1.4e-97;  
 Matches 467; Conservative 183; Mismatches 431; Indels 354; Gaps 40;  
 QY 24 QVCTGDMKRLRPASBETHLDLRLHLYGCCVQVQGNLEITVPT-NASISFLQDIQEVG 82  
 DB 100 KICIGKSRSLVSPSKNEHHYRLDRYNTCTYVDSGLKLTWLPNELDLSPDNTREVTG 159  
 QY 83 YVLIANQVRQVPLRLRIVRGTOF-----EDNYALAVLDNGDPLNNTPTVYGASPGGL 137  
 DB 160 YLISHVDVKVFPPLQITIRGTLFS.SVEEKXALFV-----TYSKM 203  
 QY 138 RELQRLSRLTEILKGVLIQNRNQLCYQDTILMKDIFPKNQALALILDTNRRACHPGSP 197  
 DB 204 YLLEIPDLNDVINGQVGFNNYNLCHMRTIQMSSEIVNSGTDAYVYDFAPRECPKHE 263  
 QY 198 MKGSGCWESSESDQSLRTVCAGGCA--RCKGFLPTCEGCAAGCTGRHSDCLAC 255  
 DB 264 SCTHG-CWGEGRKNQKFSKLTCSPOCAGRGYGRKPRBCGLFCAGGCTGTGTQDCLAC 322  
 QY 256 LHPNHSIGICELHPALVTYNTDTFSMNPBGRYTFGASCYACFYNYLSTVDSGCTLVC 315  
 DB 323 KNFFDEAVSKBECPPRKRYNPTTYVLETPEBKIVAYGATCVKCEP-GHLLRDNGCAVSC 381  
 QY 316 PLHNOVTAEDGTQCEKSKPCARVCYGLGMQYIKANSKFIGITEL-----EPAGCKX 369  
 DB 382 PODKMDKGE-----CVPNGPCPKTC-----PGVTLNAGNIDSRNCTV 422  
 QY 370 IFGSLAFPESEFDG--DPAASNTA-----PLQPEQLQVETLEETIGLYLISAWPDSLPD 421

```

Db 423 IDNIRILDTFSGFDVYVANYMTGPRYIPLDEREVEFTVEIIGYLNIEGTHPOFNR 482
Qy 422 LSYEONIQVIRGILHNHAY-STLQGLISWGLRSRLSGGLALIHNTLCEVHTV 480
Db 483 LSYFRLIETHGQWESMAALAIYSSLSYSEMRNLKISSGSGVIGNRLCYVSN 542
Qy 481 PMDQLFRNPHQALLHTANPEDECVSEGLACHOLCARGHGWGPTQCVNCSQFLRGQEC 540
Db 543 RMPAIQKEPEQKWVWNNLRADLCENKXGTCSDQCNEDGCGWAGTQCLTCKNFNNGTC 602
Qy 541 VESGRVYLGRLREYVNAHRLCPHRECOFNGSVTCGPRADCCVACAHKXDPFCVARC 600
Db 603 IADCGYISNAKY-FDNRTCKICHPCR-----TCNAGADHCQECVHVDGQHCVSEC 654
Qy 601 PFNNF-----TVS----- 608
Db 655 PKKYNDRGVCRECHATCDGCTGPKDTIGIAGCTTCLNLIINNDATYKRLDKDCPDG 714
Qy 609 -FWLRVKSASHLEPDEGACQ-----CPT-----NCTH----- 638
Db 715 YFW-----EYVHPQEGSLKPLAGRAVCRKHPLCELTNYGHEQVCSKCTHYKRR 766
Qy 639 -----SC-----VDLDDG----- 647
Db 767 EGCEBGRADHYTDEBEREGCFQRHRECGCTGPGADCCSKRFXPLDANEETGPVNSTM 826
Qy 648 -----CPAEOR-----ASPLTS-----IVSAVVGILLVVLG 674
Db 827 FNCTSKCPLEMRHNVNYQTAIGPYCAASPRSSKITANDVNNIFITIGAVLPTICILC 886
Qy 675 VVFGILIKRQOKIRKXT-MRRLQETELVEPLTSGAMPNQAOMRIKETELRVKYL 732
Db 887 VV-TYICROKQAKKKTATKALSCSESEPLRSNIGANICKRIYDABLRKGVY 944
Qy 733 GSGAGFTVYKGIWIPDGENYKIPVAIKVIRENTSPRANKIIDEAVYAGVSGPYVRL 792
Db 945 GMAFGKRVYKGVWVEGENYKIPVAIKELKSTGASESEFLEAVYIMASEHVNLLKLL 1004
Qy 793 GILLSTVQCLVTLMPGCLLDVHRENRGLSGODLLNMCQAKMSYLEDVRLVHRL 852
Db 1005 AACVMSQOMLITQMPGCLLDVHRRNRDIXGKALLNMSTQAKGSIYEEHRLVHRL 1064
Qy 853 AARNVLVKSNNHYKITDFGLARLLDIDETEHADGKVPKMMALESIILRRFTHOSDV 912
Db 1065 AARNVLVQTSVLKITDFGLAKLLSSDSNEYKAGGMPKMLALCIRNVRTSKSDW 1124
Qy 913 SYGVTWELMTFGAKPYDGIIPABEIPDLKESERLPOPTCTIDVTYIMVKCMIDSECR 972
Db 1125 AFGVTIWEIITFGQRPHEINIPADIDPLIEVGLKLPQELICSDICTLSCWHLDAAAR 1184
Qy 973 PRFRELVESESRMARDFQFVVIQNEIDL--PASPLDSTFYRSLLED--DWGDLVDAE 1027
Db 1185 PTKKQLTTVAEFARADGRYLAIPGDKFTLPA-----YTQDEKDLIRKLAFTTDS 1237
Qy 1028 EYLVPOGCFPCDPAPAGAGMVHRRSSSTRSGGDLTLGLEPSEEAR-----RSLP 1081
Db 1238 EALAKPDDYLOQPAAPSPS-----HRTDCT-----DEMPKLNRKYCKDS 1276
Qy 1082 APEBAGSDVFDG--DLGMAKAGLQSLPTHPSQLRYSBPTVPLSEDTGYAAPT 1138
Db 1277 NKMSSTGDDERDSAREVGNR-----IDLPEDEDDYLP--T 1314
Qy 1139 CSBQPEYVNOQDVAPQPSPREGFLPAPAPAGATLBRAKTLSPGKGVYKDVAFAGAVE 1198
Db 1315 CQGPNNNNNNM-----NENQNNMAVGAAGM-----DLIGVPSVD 1353
Qy 1199 NPEYL-----TPGGAAPQH-----PPAPSP-AFNDLYYWD 1230
Db 1354 NPEYLLNAQTLGVGSEPTIQTIGIPVGGPGTMEVKVPMGPSEFTSSDHEYND 1408

```

RESULT 12

```

ERBB_ALV
ID ERBB_ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN v-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=85228222; PubMed=2988784;
RX Nilgen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor."
RL Cell 41:719-726 (1985).
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M10066; AAA48763.1; ALT_INT.
DR PIR: A00643; TVCHLV.
DR PIR: B00643; TVFVLV.
DR HSP: P11362; 1FGK.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00109; TYRKINASE.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00219; TYRC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
DR Glycoprotein; Phosphorylation.
DR DOMAIN 132 399 PROTEIN KINASE.
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

Query Match 25.0%; Score 1708.5; DB 1; Length 634;
Best Local Similarity 51.5%; Pred. No. 1.6e-84;
Matches 369; Conservative 75; Mismatches 136; Indels 137; Gaps 19;

Qy 587 CAHYNDPFCVARGCF-----NFTYSPFLARVPKXSASLLEPDEBACQPCINCHSVD 642
Db 3 CAHFDGPHCVKACAGVLENDLI-VKTA-----DANAVCOLCHPNCRTGCKG 51
Qy 643 LDDKCPAEORASPLTSIVSAVY-GILLVVLGVVFGILIKRQOKIRKXTMRRLQETE 701
Db 52 PGLBSCP--NGSKTPSIAGVGGILCLVAVVGIGIYLARR-HIVKRTLRRLQERE 107
Qy 702 LVEPLTSGAMPNQAOMRIKETELRKXKVLGSAFGIVYVGIMIPDGENYKIPVAIKYL 761

```

Db	108	LVEPLTSSGEAPNQHRLIKETFEKKYKVLGSGAFGVYGLMTPEBEKKIIVALIKEL	167
Qy	762	RENTSPKANKELDEAYVMAGVSPYVSRLLIGICTSTVQLVTQMPYGCULDHVRENRG	821
Db	168	REATSPKANKELIDEAYVMAGVDPNPHVCRLLIGICTSTVQLITQIMPYGCLLDVIREHKD	227
Qy	822	RLGSODLNMCMQIAKGMSTYEDRVLYHRDLAARNVLYKSNHVKITDPGLARLLIDET	881
Db	228	NIGSQYLVNMCVCQIAKGMNLYEERLYHRDLAARNVLYKTHQYKIIDPGLAKLGADEK	287
Qy	882	EYHADGCKVPIKMMALLESILRRPFTHQSDVMSYGVYWEMLTFGAKPYDGIPAREIPDL	941
Db	288	EYHAECKGVPIKMMALLESILHRYTHQSDVMSYGVYWEMLTFGSKPYDGPASEISSEVL	347
Qy	942	EKGERLPPPICTIDVYVIMTKCMWIDSECPREPRELVSSESRMAKDRQRVVIQ-NEDL	1000
Db	348	EKGRRLPPPICTIDVYVIMTKCMWIDSPRPRELVAESKMAKDRPRLVIOGDGRM	407
Qy	1001	GPASPLDSTFFRSLIEDDDMDKDLVAEEYLVPOQGFPCPDPAFGAGVHHRRHSSSTRS	1060
Db	408	HLPSPTDSKIFRTLMEEDMEDVIDADYLVLPHQGF-----NSPST--	449
Qy	1061	GGGDLTLGLEPSEEBAPRSPF----APSEGAQSDVPDGLGMAKALQSLPRHDSPL	1115
Db	450	-----SRTPLLSSLSATSNNSNATNCID-----RMGGCHPVREDSFV	485
Qy	1116	QRYSEDPVPLPSET--DGYVAPLTSCPQPEYVNPQVPRPQPSRREBPRLAARPAKATL	1173
Db	486	QRYSDPPLGNFLSESIDDGFL-----PAPRYVNO--LMPKPS-----	521
Qy	1174	ERAKTLSPGKNGVNVDF-----AFGAIVENPEYLTPOGGAAPQHPPEPA	1218
Db	522	-----TAMVQNIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL-----XTNQSPFL	568
Qy	1219	FSPAFDNLYXNDQ-----DPPF-----RGAPSTFKGTPAENPEYLGIDVP	1260
Db	569	AKTVFESSPYWIQSGNHQINIDNPYQDDPLPNETKPGLLKVPAAENPEYLVAAAP	625
RESULT 13			
ERBB_AVIER			
ID	ERBB_AVIER	STANDARD;	PRT; 604 AA.
AC	P00535;		
DT	21-JUL-1986 (Rel. 01, Created)		
DT	21-JUL-1986 (Rel. 01, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Tyrosine-protein kinase transforming protein erbB (BC 2.7.1.112).		
GN	V-ERBB		
OS	Avian erythroblastosis virus (strain ES4).		
OC	Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.		
OX	NCBI_TaxID=79685;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=H;		
RX	MEDLINE=84026539; PubMed=6313329;		
RA	Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.,		
RT	"The erbB gene of avian erythroblastosis virus is a member of the src		
RL	gene family."		
RL	Cell 35:71-78(1983).		
RN	[2]		
RP	SEQUENCE OF 1-152 FROM N.A.		
RX	MEDLINE=84231957; PubMed=6328658;		
RA	Debnire B., Henry C., Benaisa M., Bisette G., Claverie J.-M.,		
RA	Saule S., Martin P., Stehelin D.,		
RT	"Sequencing the erba gene of avian erythroblastosis virus reveals a		
RL	new type of oncogene."		
RL	Science 224:1456-1459(1984).		
CC	-I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein		
CC	tyrosine phosphate.		
CC	-I- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND		
CC	ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS		
CC	IN CHICKENS.		

```

CC      -1- MISCELLANEOUS: V-EEB1 IS A TRUNCATED AND MUTATED VERSION OF THE
CC      RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL; K02006; AAA42394.1; ALT_INIT.
CC      DR      EMBL; K01216; AAA42400.1; -.
CC      DR      PIR; A06444; TYVUH.
CC      DR      HSSP; P11362; 1FGK.
CC      DR      InterPro; IPR000719; Euk_pkinase.
CC      DR      InterPro; IPR01245; Tyr_pkinase.
CC      DR      Pfam; PF00069; Pkinase; 1.
CC      DR      ProDom; PD000001; Euk_pkinase; 1.
CC      DR      SMART; SMO0219; TykKc; 1.
CC      DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
CC      DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
CC      DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
CC      DR      K1M transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC      K1M Glycoprotein; Phosphorylation.
CC      FT      DOMAIN 132 399 PROTEIN KINASE.
CC      FT      NP_BIND 138 146 ATP (BY SIMILARITY).
CC      FT      BINDING 165 165 ATP (BY SIMILARITY).
CC      FT      ACT_SITE 257 257 BY SIMILARITY.
CC      FT      CONFLICT 29 29 R -> W (IN REF. 2).
CC      FT      CONFLICT 140 140 S -> F (IN REF. 2).
CC      FT      CONFLICT 146 146 I -> V (IN REF. 2).
CC      SQ      SEQUENCE 604 AA; 67633 MW; 76EBD0D6745D609 CRC64;
CC
Query Match 24.5%; Score 1677; DB 1; Length 604;
Best Local Similarity 51.6%; Pred. No. 7,36-83;
Matches 361; Conservative 73; Mismatches 126; Indels 140; Gaps 18;
CC      587 CAHYKDPFPCARCP-----NFTVSFWLRPKVSASHLEPDEGACPCPINCTHSCVD 642
CC      Db      3 CAHIDGRHCYKACPAVLGENDLT-----YRKA-----DAAVQGLCPNRTCKGK 51
CC      643 LDDKGCAPEDRASPSTISAVV-GILLVVLGVVFGILIKRQOKIRKYMRLDQTE 701
CC      Db      52 PGLTGCP---NGSTPSTPIAGVVGGLCLVVGGLGILUYRR-HIYKRTLRRLQERE 107
CC      702 LVEPLTSGAMPNACNRILLKETELRVKYLGSAPFTVVKGMIPGGENKIKIVALKV 761
CC      Db      108 LVEPLTSGEAPNDAHRLILETEFKKYKVLGSGAFGIYKGLMIPREKXIVALKEL 167
CC      762 RENTSPRANKELIDEAVYMGVGSPPVYSRLGLICTSTVQLTOLMPEYGCLLDHVRENK 821
CC      Db      168 REATSPRANKELIDEAVYMASVDNPHVCRLGLICTSTVQLITGLMPYGCLLDIYREKD 227
CC      822 RLGGQDLLNMCQIAKGMSTYEDVRLVHRDLAANVTLYKSNHYKITDFGLARLLDDET 861
CC      Db      228 NIGQYLLNMCVQIAKGMNVLLEERLVRDLAANVTLYKTPQHKYITDFGLAKLIGADEK 287
CC      882 EYHADGKVPRIKMMALLESILRRRTTHOSDWSYCVTWELMTFAKPYDGIPIAREIDLL 941
CC      Db      288 EYHAEKGKVPRIKMMALLESILHRIYTHOSDWSYCVTWELMTFSSKXYDGIPIAREISSVL 347
CC      942 EKERGLPQPPICITDVTMIMKCMIDSECRPRPRELVSEFSRMAKDPQRFVYIQ-NEDL 1000
CC      Db      348 EKGGRLLPQPPICITDVTMIMKCMIMADSRPKFERELIAEFSKVAKDPFRLYVQSGERM 407
CC      1001 GPASPLSTFRSLLEDODGDIVDAEYLVYQGGFCPCPDPAAGAMVHHRRSSSTRS 1060
CC      Db      408 HLPSPFDSKFRILMEEDMEDIVDAEYLVYHVGGF-----NSPFT- 449
CC      1061 GGGGLTLGLEPSEBEARSP-----APSEGAGSVDGDLGMGAAGLOSLPTHDSPL 1115

```





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:22:34 ; Search time 48.5887 seconds

(without alignments)  
5347.444 Million cell updates/sec

Title: SEQ4-579-593-14

Perfect score: 6841  
Sequence: 1 MELALCRMGLLLALLPAGA.....TFKGTPTAENPEYLGHDPV 1261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SPTREMBL\_21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mnc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriophage.\*
- 17: sp\_archaea.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	6114	89.4	1259	6	Q18735
2	3099	45.3	1209	11	Q9QX70
3	3077	45.0	1210	11	Q9EP98
4	2687	39.3	1165	13	Q9YH40
5	2667.5	39.0	1137	13	Q9W6F6
6	2259	33.0	1328	13	P79754
7	2026.5	29.6	1433	5	Q9B1H9
8	1871	27.3	419	4	Q9UK79
9	1739	25.4	367	11	Q8R2X1
10	1697.5	24.8	412	4	Q8WYV0
11	1679	24.5	729	15	Q86712
12	1677	24.5	567	15	Q86714
13	1612.5	23.6	962	15	Q84895
14	1604	23.4	545	15	Q85468
15	1468.5	21.5	655	11	Q9WVFS
16	1452.5	21.2	643	11	Q9ERV6

17	1273	18.6	1193	5	Q9Y1X8
18	1206.5	17.6	1368	5	Q23821
19	1176	17.2	1717	5	Q26566
20	1126	16.5	527	13	Q90836
21	1001.5	14.6	478	11	Q9E8E0
22	906	13.2	165	4	Q14256
23	901.5	13.2	599	13	Q9PSH2
24	806.5	11.8	346	13	P11776
25	783	11.4	176	11	Q923V5
26	778	11.0	435	5	Q8SZM1
27	754.5	11.0	311	13	Q99162
28	734	10.7	331	4	Q9EUD7
29	731	10.7	1671	5	Q9N0V5
30	723	10.6	149	6	Q9B6G6
31	720.5	10.5	1362	13	Q9PV24
32	693	10.1	1418	13	Q93457
33	685.5	10.0	1368	13	Q8UW85
34	670.5	9.8	1369	13	Q8UW86
35	668.5	9.8	1472	5	Q9U5A8
36	665	9.7	1358	13	Q73798
37	653	9.5	1412	13	Q8UW84
38	644.5	9.4	1418	13	Q8UW83
39	633	9.3	1245	13	Q9YGH8
40	628	9.2	1371	11	Q9QVW4
41	619.5	9.1	2144	5	Q9YD94
42	592	8.7	987	11	Q91YMO
43	587.5	8.6	1036	4	Q07912
44	587.5	8.6	1055	11	Q05967
45	587	8.6	935	4	Q96J35

## ALIGNMENTS

RESULT 1

ID 018735 PRELIMINARY; PRT; 1259 AA.

AC 018735;

DT 01-JAN-1998 (TEMBUREL. 05, Created)

DT 01-JAN-1998 (TEMBUREL. 05, Last sequence update)

DT 01-JUN-2002 (TEMBUREL. 21, Last annotation update)

DE ErbB-2.

OS Canis familiaris (Dog).

CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI\_Taxid=9615;

RN [1]

RP SEQUENCE FROM N.A.

RA Yokota H.;

RT "cDNA cloning of erbB-2 from canine mammary gland";

RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.

DR EMBL; AB008451; BAA23127.1; -

DR HSSP; P1362; 1FEK.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000494; EGFR\_L domain.

DR InterPro; IPR000719; Euk\_DK\_kinase.

DR InterPro; IPR002174; Purih-like.

DR InterPro; IPR001245; Tyr\_kinase.

DR InterPro; IPR004019; YLP\_motif.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; Pkinase; 1.

DR Pfam; PF01030; Recep\_L\_domain; 2.

DR Pfam; PF02757; YLP\_2.

DR ProDom; PD000001; Euk\_Pkinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00219; TyKc; 1.

DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_1.

DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE; PS0011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE; PS0109; PROTEIN\_KINASE\_TYR; 1.

KW ATP-binding; Transferase; Tyrosine-protein kinase.

SEQUENCE 1259 AA; 137989 KW; E37364D49C4ACD46 CRC64;

Query Match 89.4%; Score 6114; DB 6; Length 1259;  
 Best Local Similarity 88.7%; Pred. No. 0;  
 Matches 1133; Conservative 44; Mismatches 66; Indels 34; Gaps 4;

QY 1 MELALCNGWLLALLPGCASTQVCTGDMKRLPASPETHLDMRLHLYQCCQVQVGNL 60  
 DB 1 MELALCNGWLLALLPGCASTQVCTGDMKRLPASPETHLDMRLHLYQCCQVQVGNL 60  
 QY 61 ELTYLPTNALSFLDIOEVOGVYLIANQVAVPRLRIRGTQLFEDNVALAVLNG 120  
 DB 61 ELTYLPTNALSFLDIOEVOGVYLIANQVAVPRLRIRGTQLFEDNVALAVLNG 120  
 QY 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPQLCYDTIIMKDIHKNOLA 180  
 DB 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPQLCYDTIIMKDIHKNOLA 180  
 QY 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPQLCYDTIIMKDIHKNOLA 180  
 DB 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGVLIQRNPQLCYDTIIMKDIHKNOLA 180  
 QY 181 LTLIDNTRRACHPCSPMCKSGRCWSESEDCQSLTRYCAGGACRCGRLPTDCCHQC 240  
 DB 181 LTLIDNTRRACHPCSPMCKSGRCWSESEDCQSLTRYCAGGACRCGRLPTDCCHQC 240  
 QY 241 AAGCTGPKHSDCLAHFNHSGICEIHCALVTYNTDTFESMNPGRYTFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLAHFNHSGICEIHCALVTYNTDTFESMNPGRYTFGASCVTACP 300  
 QY 301 NYLSTDVSGCTLVCPHANOBTADGTORCKECSKPCARVYCYGMOTIKANSKFTIGT 360  
 DB 301 NYLSTDVSGCTLVCPHANOBTADGTORCKECSKPCARVYCYGMOTIKANSKFTIGT 360  
 QY 361 ELEFAGCKIFGSLAFLEPSFDGDPASNTAPLOPQLOVFEITGLYLSAMPDGLP 420  
 DB 361 ELEFAGCKIFGSLAFLEPSFDGDPASNTAPLOPQLOVFEITGLYLSAMPDGLP 420  
 QY 421 DUSVTONLOVIRGRILHNGAVSLTQGLISWLGRLSRLSGLALIHNTLHLCFVATV 480  
 DB 421 DUSVTONLOVIRGRILHNGAVSLTQGLISWLGRLSRLSGLALIHNTLHLCFVATV 480  
 QY 481 PMDOLFRNHQALLTANRPEDECEGEGACHQCLARHGWPGSTOCVNSQFRRGEC 540  
 DB 481 PMDOLFRNHQALLTANRPEDECEGEGACHQCLARHGWPGSTOCVNSQFRRGEC 540  
 QY 541 VEECRVLOGLPREYVNAHRLCPHRECPONGSVTCFGEADQCVACAHYKDPFVCARC 600  
 DB 541 VEECRVLOGLPREYVNAHRLCPHRECPONGSVTCFGEADQCVACAHYKDPFVCARC 600  
 QY 601 PENNFTVSFWLEVPKVASHLEP-----DEBACQCPINCTHSCVDLDDKCP 649  
 DB 601 PENNFTVSFWLEVPKVASHLEP-----DEBACQCPINCTHSCVDLDDKCP 649  
 QY 650 AEORASPLTSIVSAVVGILLVVLGVFGLIKRQCKIRKTYMRLLOETELVEPLTPS 709  
 DB 650 AEORASPLTSIVSAVVGILLVVLGVFGLIKRQCKIRKTYMRLLOETELVEPLTPS 709  
 QY 710 GAMPOAOQRIILKELEIKVYKVLGSGAGTGYKGIWIDGENVKIPIVAKYLRENTSFA 769  
 DB 710 GAMPOAOQRIILKELEIKVYKVLGSGAGTGYKGIWIDGENVKIPIVAKYLRENTSFA 769  
 QY 770 NKEILDEAVYVAGVSPYVRLGICLTSTVOLVQLMPYGLDLHVENRGRLGSOGL 829  
 DB 770 NKEILDEAVYVAGVSPYVRLGICLTSTVOLVQLMPYGLDLHVENRGRLGSOGL 829  
 QY 830 NWCQOIAKGMSTLEBVLVHRDLAARNVLYVSPNHVKTDFGLARLLDIDETEVHADGK 889  
 DB 830 NWCQOIAKGMSTLEBVLVHRDLAARNVLYVSPNHVKTDFGLARLLDIDETEVHADGK 889  
 QY 890 VPIKMMLESILRRRFTHOSQVYSGVTVMELMTGAPRYGIPARETDLLEGEPLPQ 949  
 DB 890 VPIKMMLESILRRRFTHOSQVYSGVTVMELMTGAPRYGIPARETDLLEGEPLPQ 949  
 QY 950 PPICTIDVYMIWVKCMTIDSECRPFRELVEFSFMSADPQRFVVIQNEDEDGSPASPLDST 1009  
 DB 950 PPICTIDVYMIWVKCMTIDSECRPFRELVEFSFMSADPQRFVVIQNEDEDGSPASPLDST 1009  
 QY 1010 FYRSLLEDDMDGLVDABEVLVPOQGFPCPDAPAGMWHHRSSSTRSGGDLTLGL 1069  
 DB 1010 FYRSLLEDDMDGLVDABEVLVPOQGFPCPDAPAGMWHHRSSSTRSGGDLTLGL 1069

DB 1003 FYRSLLEDDMDGLVDABEVLVPOQGFPCPDAPAGMWHHRSSSTRSGGDLTLGL 1062  
 QY 1070 EPSEEARSPPLABSGAGSDVDGDIQWGAAGLQSLPHDPSPLORYSEDPVLPSE 1129  
 DB 1063 EPSEEARSPPLABSGAGSDVDGDIQWGAAGLQSLPHDPSPLORYSEDPVLPSE 1122  
 QY 1130 TDGVVAPLTCSPQPEYVNOQDVPQPPSPREGPLPARPAGATLER-----AKTISPKN 1184  
 DB 1123 TDGVVAPLTCSPQPEYVNOQDVPQPPSPREGPLPARPAGATLER-----AKTISPKN 1182  
 QY 1185 GVXVDVAFGAVENPEYLTPOGGAAPQPPPPAFSPAFDNLVYWDQDPERGAPESTFX 1244  
 DB 1183 GVXVDVAFGAVENPEYLTPOGGAAPQPPPPAFSPAFDNLVYWDQDPERGAPESTFX 1242  
 QY 1245 GTPTAENPEYLGIDVPV 1261  
 DB 1243 GTPTAENPEYLGIDVPV 1259

## RESULT 2

Q9QX70 PRELIMINARY; PRT; 1209 AA.  
 AC Q9QX70;  
 DT 01-MAY-2000 (TREMBlrel. 13, Created)  
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor.  
 GN EGFR.  
 OS Rattus norvegicus (Rat).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JFISHER; TISSUE=LIVER;  
 RX MEDLINE=30236886; PubMed=2342466;  
 RA Petch L.A.; Harris J.; Raymond V.W.; Blasband A.J.; Lee D.C.;  
 RA Earp H.S.;  
 RT "A truncated, secreted form of the epidermal growth factor receptor is  
 RT encoded by an alternatively spliced transcript in normal rat tissue";  
 RL Mol. Cell. Biol. 10:2973-2982(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JFISHER; TISSUE=LIVER;  
 RA Petch L.A.;  
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=JFISHER; TISSUE=LIVER;  
 RA Guttridge K.; Dawson T.L.; Earp H.S.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: M37394; AAF14008.1; -.  
 DR HSSP: P1362; IFCK.  
 DR InterPro: IPR000494; EGFR\_L domain.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002174; Euk\_Pkinase.  
 DR InterPro: IPR001245; Tyr\_Pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF01030; Recept\_L domain; 2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00261; FV; 3.  
 DR SMART: SM00219; Tyrc; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE\_ATP; 1.  
 DR PROSITE: PS00111; PROTEIN KINASE\_DOM; 1.  
 DR PROSITE: PS00109; PROTEIN KINASE\_TYR; 1.  
 KM APP-binding; Receptor; Transferrase; Tyrosine-protein kinase.  
 SQ SEQUENCE 1209 AA; 134891 MW; 96FE7F6CIB7773 CRC64;

Query Match 45.3%; Score 3099; DB 11; Length 1209;  
 Best Local Similarity 49.9%; Pred. No. 1; 7e-225;



Matches 643; Conservative 161; Mismatches 353; Indels 132; Gaps 31;

QY 3 LAALCRWGLLALLPSCA-ASTOYCTGTDMKRLPASPETHLMLRLHYOGCVOGNLE 61  
 Db 15 LAALCAAG-----GALBEKKVCGSTNRLTGLTFEDHPLSLQMFNNCEVVLGNLE 66  
 QY 62 LTLPLTASISLQDIOVOGYVLIANQVRQVPLQRLRVGTGLFEDNYALAVDNGD 121  
 Db 67 ITTVORNYDLSPLKLTQEVAGYVLIANTVERIPLNLTQIRNATYENTYALAVLSN- 124  
 QY 122 PLNNTPVTGASPGSLRELOLRSITLILKGVLIQRPOLCYODTILMKDIFKNNQLAL 181  
 Db 125 -----YGTNNKGLRELPMRNLQELLIGAVRFSNNPILCMNETTQMRTIV-QDPLSN 175  
 QY 182 TLIDTNRS-RACPCSCMCCKGSCRCWSSSDCCSLRTVACGCA-RCKGPLPTDCHEQ 239  
 Db 176 MSMDVQRHLTGCPKDCSPNGSCMGSGEENCKLTXTICAQOCCSRRCRGRSPDCCHQ 235  
 QY 240 CAAGCTGPKASDCIACIHFHNSGICELHCPALVTYNTDPESMPNEGRTFGASCVTAC 239  
 Db 236 CAAGCTGPRSDCLVCHRFDEATCKTCCPLMLYNPTTQMNVNBEKISFGATCVKXC 295  
 QY 300 PNYVLTSSCTLVCPHNOEYTAEDGTGRCCKSKPCARVCYGLGMOYIKANSKFIGI 359  
 Db 296 PRNVVYTDHSGCVARACGPDYEV-EEDGVSKCKKCKDOPCKKVCNG:GIGEFK-DTLSTNA 353  
 QY 360 TELE-FAGCKKIFGSLAFLEPSFDGPASTAPLOEQLOVFTTEITGYLYISAMPDS 418  
 Db 354 TNRKHFYCTAIGSDHLFLVARKGSFTSTPPLDPELILTVETLVEITGFLIQAMPEN 413  
 QY 419 LPDLVFNQLVIRGRILHNGAVSLTLOGISMLGRSLREKSGSGLALHNTHLCFYH 478  
 Db 414 WTLDAFENLEIRGRKQHGQSLAVVGLNITSLGRSLKESIDSDVILISGRNLCYAN 473  
 QY 479 TVWMDQLFRNPHQALHTANRPEDECVGSLACHQICANGCHGCPPTQCVNCSQFLRQ 538  
 Db 474 TIMWKLFGTPNOKTKIMNRAKCKATNVCNPLCSSGSGCEPPTDCCVSCQNSRGR 533  
 QY 539 ECVCEGRVLQGLFREYVNAHNCPCHECOPONGSVTCSEPEADCCVCAHYKXDPFCVA 598  
 Db 534 ECVDKCNILEGEPREVENECICQCHPECLPQMNITTCGRGDNCTCKAHYVDGHCVK 593  
 QY 599 RCFP-----NFTVSWLRVYKVASHLBDEEGACPCPINTHSCVDLDDKGPAPOR 653  
 Db 594 TCSGIMGENNTLV-W-----KFA-----DANNVCHLGHANTCYCAGBGKGC--QGP 639  
 QY 654 ASP-LTSYSAVVGIILVVLGVVFGI-LIKRQCKIRKXTMRLLQETLVEPLTPSCA 711  
 Db 640 EGKIPSIATGIVGLLFTIV-VALGIGLFMRRLQVRRKTLRLQLQERELEVEPLTPSGE 698  
 QY 712 MPNOQMRLIKETELRKVYLGSAFGTYKGIWIDGENVKIPVAIKVLRNTSPKANK 771  
 Db 699 APQQAHLRIKELKEFKKIKVLSGAGFYKGLIMEGEKVIKPAIKLRENTSPKANK 758  
 QY 772 EILDEAYVNAVGVSRYVSRLLGLCTSTVQLVLTQMPRYCQLLDHVENRGRISQDNLW 831  
 Db 759 EILDEAYVNAVDNPHVCRLGLCTSTVQLVLTQMPRYCQLLDVYEHKNDIGSYQLNWM 818  
 QY 832 CMQIAGKMSYLEDVRLVHRDLAARNVLVKSPNVKTTDGLARLLDIDETVHAOGKRP 891  
 Db 819 CVOIAGKMYLEBRRLVHRDLAARNVLVKTPQVKTDDGLAKLCAEKEKHABGKRP 878  
 QY 892 IKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPOP 951  
 Db 879 IKMALESILHRYTHQSDVMSYGVTVWELMTFGSKPYDGIAPASESSILKGERLPOP 938  
 QY 952 ICTIDVYMTVMKCMIDSECPREFRLVSEFSRMAPDRFVYIQ-NEULGASPLDSTF 1010  
 Db 939 ICTIDVYMTVMKCMIDSDSRPFRLVSEFSRMAPDRFVYLVIGQDERMHLPSPTDSNF 998  
 QY 1011 YRSLLEDDMDGLVDAAEYLVDPQGFCDPAPGAGVHHHRSSSTFSGGDITLIGLE 1070  
 Db 999 YRALMEEDMEDVDADEVLTLPQGGF-----NSPST----- 1030

QY 1071 PEEERAPRSPLAPSEGAGSDVFDGLGMAKGLQSLPTHDPSPQRYSEDPVLPSET 1130  
 Db 1031 -----SRPFLSLISANSN-----SSTVACIRNRNSCKVKEAFQRIYSDDPTSVLTEN 1080  
 QY 1131 --DGYVAPLTGSPQPEYVNPQVPRQPSPREGPIPAARPAATLIERAKTSPGXGVK 1188  
 Db 1081 IDDTFL-----FVPEYINQ-SVPRKPAQSVQNPVYHNPFLHP-----APGRDLHYQ 1125  
 QY 1189 DVFAFGAGAVENPEYL-TPQCGAAPQHPHPAFSPAFDNLVYWDQ-----DP----- 1233  
 Db 1126 N--PHSNVANSREYLTAC-----PTCLSSGFDSSALMTQKGSQMSLDNPDYQD 1174  
 QY 1234 --PERGAPSTFKGTPTAENPEYLGIDVP 1260  
 Db 1175 FEPKPAKPNQIFKG-PTAENPEYLRVAP 1202

RESULT 3  
 ID Q9EP98 PRELIMINARY, PRT, 1210 AA.  
 AC Q9EP98;  
 DT 01-MAR-2001 (TREMblrel. 16, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor isoform 1.  
 GN EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid:10090;  
 RN [1]  
 RC SEQUENCE FROM N. A.  
 RP STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 RA Sinclair C.S., Pearall R.S., Green P.J., Yee D., Lampand A.L.,  
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
 RA Maile N.J.;  
 RA "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative Egr transcripts encoding truncated receptor  
 RT isoforms."  
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RC SEQUENCE FROM N. A.  
 RP STRAIN=C57BL/6J;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 RA Schell C., Pearall R.S., Green P.J., Yee D., Lampand A.L.,  
 RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
 RA Maile N.J.;  
 RA "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative Egr transcripts encoding truncated receptor  
 RT isoforms."  
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AF275366; AAC28045.1; -  
 DR EMBL; AF275364; AAC28045.1; JOINED.  
 DR EMBL; AF275365; AAC28045.1; JOINED.  
 DR EMBL; AF275367; AAC24386.1; -  
 DR HSRP; P11362; IRK.  
 DR MGD; MGI:95294; Egfr.  
 DR InterPro; IPR000345; Cytochrome b5.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR InterPro; IPR000719; Euk\_Dkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR002290; Ser-THR\_Dkinase.  
 DR InterPro; IPR001245; Tyr\_Dkinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Kinase; 1.  
 DR Pfam; PF01030; Recep\_L domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_Dkinase; 1.  
 DR SMART; SM00261; FU; 5.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR SMART; SM00219; Tyrc; 1.

DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN\_1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP\_1.  
 DR PROSITE: PS00111; PROTEIN KINASE\_DOM\_1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR\_1.  
 DR ATP-binding; Receptor; Transferase.  
 SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9D3E2E18 CRC64;

Query Match 45.0%; Score 3077; DB 11; Length 1210;  
 Best Local Similarity 49.5%; Pred. No. 7,88-224;  
 Matches 635; Conservative 160; Mismatches 361; Indels 128; Gaps 27;

QY 11 LLLALLPFGAA-STVCTGTDMKRLPASPETHLDMRLHLYGCGVQVGNLELYLPTN 68  
 DB 14 LITLALCAAGALAEKKYCOGTSNRLTDLGTDFEDHFLSLQRMYNCEVLANLELTYQRN 73  
 QY 69 ASLSELDIOIEVGVVLIHNVQVLPQRLIRYRGVQLFEDNVALVLNGLDPLNNTTP 128  
 DB 74 YDLSPFLKIOEVAGVLIHNVQVLPQRLIRYRGVQLFEDNVALVLNGLDPLNNTTP 124  
 QY 129 VTGASPGGLRELQLSLTEILKGVLTQRNPOLCYODTILMKDI---FKNNQLATLLI 184  
 DB 125 -YGTNRITGLRELPRNLQELIGAVRFSNNPILCNMDTILQWRDVGWFMNSMDL--- 180  
 QY 185 DTNSRACHPCSPMKSGSRGWGSESDCCSLTRTVCAAGCA-RKGLPLPTDCCHEGCAAG 243  
 DB 181 -QSHPSGCPKCDPSCGPGSCWGGEGENCQKLTKLICQAQCSHRGRGSPDCCNQCAG 239  
 QY 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMPREGRYTFGASCVTACPYNY 303  
 DB 240 CTGPRESDCLVCCQFQDEATCKDCTPMLNPTTYQMDVNPBESKVFQATCYCKCRNY 239  
 QY 304 LSTVGSCITLVCPILHNOEVTADGTQCEKSKCAVCGGLMOYIKANSKFIQTELE 363  
 DB 300 VVTHGSCVRAQGDYEV-BEDGIRKCKCKDGCRCVKNIGIGEEK-TLISINATNIK 357  
 QY 364 -FAGCKIFGSLAFIPESFDGDPASNTAPLOPELOVEFTLEETGVLYISAMPDSLPL 422  
 DB 358 HFXYCTAISGLHLPLPAFKDSTRTPLDPRELEILKTYKELTGLLIQAMDNDTDL 417  
 QY 423 SVFQNLQVIRGRILHNGAYSLTLOGLSIWGLRSLELSGLALIHNTLHFLVHTVPW 482  
 DB 418 HAFENLEIRGTRQHGQFSLAVVGLNITSGLRSLKEISDGVIIISGNRLCYANTINM 477  
 QY 483 DQLEFRNHQALHIANRPEDECEBGLACHQLCARHGMWGPVQCYNCGQLRGQCEVE 542  
 DB 478 KKLFTGTPQKTKIMNNAEKDCXAVNHVCVPLCSSECGWGPEDCVSCONVARGRECE 537  
 QY 543 ECRVLOGLPREYVNAHQLPCHEPCOPONGSVTCFGEADQCVCAHYKDPPEVCARCP 602  
 DB 538 KCNILEGPREFEVENSEICIQHPECLPQAMNITCTGRGPNCTICAHYIDGPHCVKCPA 557  
 QY 603 -----NFTVSFMLRVKVASASHLEPDEBGCQCPINCHSCYDLDKCCPAEQRASPL 657  
 DB 598 GIMGENNTLV-WKYA-----DANNVCHLCHANCTYCGCPGLQCEWFGSPKXI 645  
 QY 658 TSISAVAGILLVVLGVVFGI-LIKRQCKIRKYTRRLIQLTELETPSGAMPNQA 716  
 DB 646 PSINATGIVGLLFIIV-VALDIGLPMRRHIVKRTIRRLIQEHELVEPLPSGEAPNQA 704  
 QY 717 QMRILKTELKAVYLSSGAFTYKGIWIPDGENVKI PAIAIKYLRNTPSPKANKETLDE 776  
 DB 705 HLRIKTEFEFKKIVLSSGAFGVYKGIWIPBEGEKVI PAIKELREATSPKANKETLDE 764  
 QY 777 AYVWAGVSPYVSRLLGLTSTVQLTOLMPYCCLDHYRENNRGRGSDLLMNCQIA 836  
 DB 765 AYVWASVDNPHVCLLGLTSTVQLTOLMPYCCLDHYRENNRGRGSDLLMNCQIA 824  
 QY 837 KQMSYLEDVLRHDLAARNVLKSPHVKITPFGARLLDIDETEHADGKVPKIKMA 896  
 DB 825 KQMYLEDRLVHDLAARNVLKTPQHVKITPFGAKLGAEEKEVHAEGKVPKIKMA 884  
 QY 897 LESTLRRTFHQSQVMSYGVTVWELMTFGAKPYDGIARELPDLLEGEELLPQPICTID 956  
 DB 897 LESTLRRTFHQSQVMSYGVTVWELMTFGAKPYDGIARELPDLLEGEELLPQPICTID 956

DB 885 LESTLRRTFHQSQVMSYGVTVWELMTFGSKRXYDGIASDISILEKGERLPQPICTID 944  
 QY 957 YVIMWKCMWIDSECRPRRELVESESRAPQPFVYIQ-NEDLGPASPLSTFRRSL 1015  
 DB 945 YVIMWKCMWIDADSPPKRELITLSESKARDPQRYLYIQGERHMLPSPTSNFRALM 1004  
 QY 1016 EDDMDGLDAEAEVLPQCGFCPCPDPAQAGMHHRRSSSTRSGGGLTLGLEPSEE 1075  
 DB 1005 DEEDMDVDVADAEYLTPQGF-----NSPST----- 1031  
 QY 1076 APPSLPSEAGSDVFDGLGMAKGLQSLPTDPSFLQYSEDPTPLPSET--DGY 1133  
 DB 1032 -SRPLSLLSATSN-----NSTVACINRNGSCRVKEDAFLQRYSSDPTGAVTEDNIDAF 1086  
 QY 1134 VAPLTCQPEEVNQPDPVPPSPREGPLPAPRAGATLEAKTLSPGKGVVDVFAF 1193  
 DB 1087 L-----PVPEVYNQ-SVXKRAGSVQNPVYNQPLHP-----APGRDHYQV--PH 1129  
 QY 1194 GGAVENPEYL-TPQGAAPQPPAPPAFPDNLVYWDQ-----DP-----PER 1236  
 DB 1130 SNAVGNPEYLNTAQ-----PTCLSGGFSPALMTQKSGHQMSLDNPVYQDPFPKE 1180  
 QY 1237 GAPSTFKGTPTAENPEYIGLDVP 1260  
 DB 1181 TRPNGIFKG-PTAENAEYLRLVAP 1203

## RESULT 4

QYH40 PRELIMINARY; FRT; 1165 AA.  
 AC QYH40;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Receptor tyrosine kinase proto-oncogene.  
 GN XMRK.  
 OS Xiphophorus xiphidium.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
 NX [1] Taxid=8086;  
 RN [1] SEQUENCE FROM N.A.  
 RP STRAIN=RIO PURIFICATION;  
 RC MEDLINE=98241172; PubMed=982016;  
 RA Dimatilevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,  
 RA Altschmidt J., Scharl M.;  
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by  
 RT overexpression and mutational alterations";  
 RL Oncogene 16:1681-1690 (1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RA Scharl M.;  
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: U53471; AAD0500.2; -;  
 DR HSSP: P11362; 1FGK.  
 DR InterPro: IPR000345; CytC\_heme\_bind.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001899; Gram\_pos\_anchor.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_kinase; 1.  
 DR SMART: SM00261; Fu; 3.  
 DR SMART: SM00219; Tyrc; 1.  
 DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN\_1.  
 DR PROSITE: PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.

DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00011; PROTEIN KINASE DOM; 2.  
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.  
 KW ATP-binding; Kinase; Tyrosine-protein kinase.  
 SQ SEQUENCE 1165 AA; 129614 MW; 7FEE38D877A74E CRC64;

Query Match 39.3%; Score 2687; DB 13; Length 1165;  
 Best Local Similarity 45.1%; Pred. No. 2,5e-194;  
 Matches 579; Conservative 161; Mismatches 385; Indels 160; Gaps 32;

QY 1 MELALCRWGLLALLPPG-AAST---CVCTGTDKMLRLPASPEHMLDLRLHLYGQGV 55  
 4 LELLEL-----LTLILSLGRCSSTDPDRKVCQGTSTNQMTW---LDNHVLYKMKMYGSCNV 56  
 QY 56 VQGNLELTYPTAASFLDLOIEVQGVYLIANNQVPLQRLVRGTQLFEDNYALA 115  
 57 VLENLETTYQENQDSFLQSIQEVGYVLIANNEVSTIPLVRLIRIQNYIEGFTLL 116  
 QY 116 VLDNGDPLNNTTAVTGASPGGLRELQRLSTELIKGVLIOENPOLCYODTLIMKDIFFK 175  
 117 VMGNVCK-NPSSP--DYVQVGLKQLQSLNLTSLSGGVKSHNPLICNVETIMWDIVX 173  
 QY 176 NNGLALLTLDTNSTRACHPCSPYCKGSRGCESSSEDQSLRTVCAGG-ARCKGLPTD 234  
 174 TSNPTWMLIPHAPEKCCOKCDPGCVNGSCWAPGPGHCQKFTKLLCAEQCNRRRGPKPID 233  
 QY 235 CCEQCAAGCTGPKHSDCLAFHNSGICELHCPALVTYNTDFESMNPREGRTFGAS 294  
 234 CNEHCAGGCTGRATDCLACRDNDGCTCKTPPKYIDVSHQVNNPINKYTFGA 293  
 QY 295 CTTACPYNYLSTDVGSCTVCPHNOEYTAEDGTQCEKSCPKARVCYGLGM---QYI 350  
 294 CKECPNSYVTE-GACVRSKSGMLEVD-ENGKRSCKRCDDGCPKVCIDIGLSNTI 351  
 QY 351 KANSKRTIGTELEFPACKKIFGSLAFLPSPGDDPKSNAPLQPELOVFELEETGYL 410  
 352 ANNSTNIG---SFSNCTKINGDIILNRSFECDDPKYKIPMDPELMLTITVKEITGYL 407  
 QY 411 YISAMPDLPDISVFQNLQVIRGRIHNGAYS-LTQGLGIMLGRSLRETSGLALIH 469  
 408 VIMMPENNSTLSVFNELIIRGRTTFSGFSFVYVQVSHQMLGSLISKEYSAGNVILK 467  
 QY 470 HNTHLCTFTVPMQDLFNNPHQALLTANRPEDECVEGLACHQDLCARGCHWPGPTQCV 529  
 468 NTPOLRYASTINMRILFRSEDOSEIYDART-----ENQCNNECEDSGCGPPTCV 520  
 QY 530 NCSQFLRGQECVEECVGLGLPREYVNAHCLPCHPECCQPGQSVTCRPPADQVACAH 589  
 521 SLIHYDRGRCVASCULQGEPRACVDRGVQCHOELVQDSTLCYGPGRANCSKCAH 580  
 QY 590 YKDPPEFCVACRP---FNNFTVSFWLVPKVSASHLEPDEGACQCPINICTHSQVLDLCK 646  
 581 FQDGFQCI PRCPHGMGLDGLTLMKYA-----DMGQCQCHQNCCTGCGGPGIS 630  
 QY 647 GCPAQRASPLTISYSAVYGLLVVYGVVFGILIKRQOKRKTMYRRLLOETELVPL 706  
 631 GGRGD-IYSHSLAVGLVSGLLITVALIVLLRRRIK-RKTIIRRLQEKELVPL 688  
 QY 707 TFSGAMPNQAQWRIKETELRKVYLGSGAFGTVYKGIWIPGGEVVKIPIVAKVRENTS 766  
 689 TFSGAPNQAQFLRIKETEFKKDRVLGSAFCTVYKGLMNPGENIIRIPVAKVRENTS 748  
 QY 767 PRANKELIDEATVMAVGSPPYSLGLICTISTVOLVQLMPYGLLDHVENRGRLSQ 826  
 749 PRVNOEVLDEATVMAVSDHPVCRLLIGICTLSAVQLVQLMPYGLLDHVENRGRLSQ 808  
 QY 827 DLNMCQIAKMSYLEDVRLVHRLAARNVLYKSPNHYKTIIDELALLDIDEYIAD 886  
 809 WLNMCQIAKMSYLEDVRLVHRLAARNVLYKSPNHYKTIIDELALLDIDEYIAD 868  
 QY 887 GSKVPIKMALESILRRFTHOSDVSYGVTVWELMTGAKRYDGIIPAREIPDLLEKER 946  
 869 GSKVPIKMALESILQMTYTHOSDVSYGVTVWELMTGSKRYDGIIPAKELASVLENER 928  
 DB

QY 947 LPQPICTIDVYIMVWKCMIDSECRPRFELVSEFSHAPDQRFVYIOWEDLGPASPL 1006  
 929 LPQPICTIEVYMIILKCMIDPSSRPFRELVSFSGMADPSYVLIQ---NLPSFS 985  
 QY 1007 DSTFYSLLIEDDMGLDYDAEYVLPQGFPCPDPAAGGVHRRSSSTRSGGDLT 1066  
 986 DRLPSRLISSD--DVYDADEYLL-----RYKRIN-RQGS----- 1018  
 QY 1067 LGLPSEEEAPRSPAPSEGAGSDVFDGDLGMAKGLQSLPTHDPSSLQSYSDPTV-P 1125  
 1019 -----EPILPENG-----PVRENSIALYISDPTQNA 1046  
 QY 1126 LPSETDGYVAPLTCSPOPEYVNPDPVREP-----PSPRE-----GPLP-AARPAAT 1172  
 1047 LEKDDGH-----EYVNPQSESTRSLDIYNNYEDLTGQGVSLSSGEATN 1096  
 QY 1173 LERAKTLPKNGVGVKDYAFAGAVENREYLTPOGGAAPQTHPPAPSPADNLYWQD 1232  
 1097 FSRPEYLTNQNLSL--PLVSSGSMDDEY---QAG-----YQAAF----- 1131  
 QY 1233 PPERGAPSTFKGTPTAENPEYLG 1257  
 1132 LPQTGALTNGMFLPAAENLEYLG 1156  
 DB

## RESULT 5

Q9W6F6 PRELIMINARY; PRT; 1137 AA.  
 AC Q9W6F6;  
 DT 01-NOV-1999 (Trembl) 12, Created  
 DT 01-NOV-1999 (Trembl) 12, Last sequence update  
 DT 01-UN-2002 (Trembl) 21, Last annotation update  
 DE Receptor tyrosine kinase (Fragment).  
 GN BRB4.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HINDBRAIN.  
 RX MEDLINE=99263203; PubMed=10328884;  
 RA Dixon M., Lumsden A.;  
 RT "Distribution of neuregulin-1 (nrg1) and erb4 transcripts in  
 RT embryonic chick hindbrain."  
 RL Mol. Cell. Neurosci. 13:237-258 (1999).  
 DR EMBL: AF121963; AAD31764.1; -;  
 DR HSSP: P11362; IFGK  
 DR InterPro: IPR000494; EGFR\_L domain.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001368; TNFR\_C6.  
 DR InterPro: IPR001245; Tyr\_Pkinase.  
 DR InterPro: IPR004019; YLP motif.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF01030; Recep\_L domain; 1.  
 DR Pfam: PF02757; YLP; 2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_Pkinase; 1.  
 DR SMART: SM00261; Fv; 3.  
 DR SMART: SM00219; TyKc; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.  
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.  
 KW Kinase; Tyrosine-protein kinase.  
 FT NON TER 1  
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match

39.0%; Score 2667.5; DB 13; Length 1137;

Best Local Similarity 46.2%; Pred. No. 7.2e-193;  
Matches 532; Conservative 171; Mismatches 353; Indels 95; Gaps 27;

QY 161 LCYQDTILMKDIFHKNNQALTLIDTNRSAHPGSPCKSGRCGSESSDQSLTRIVC 220  
DB 3 LCFADTLIMQDIDVRNPMASNFTLVPTNGSSGGRCHKCTG-RCMGPTENHQCOTLTKIVC 61  
QY 221 AGGC-ARCGPLPTDCCHEGCAAGCTGPGSDCLCLHNSGICELHCPALVTYNTDTF 279  
DB 62 AEQCDGRCYGPVSDCCHEGCAAGCTGPGSDCLCLHNSGICELHCPALVTYNTDTF 121  
QY 280 ESMRPEGRYTFGASCTVACPNYVSTDVGSCTLVCPHNOETADGTQRCCKSKPCA 339  
DB 122 QLEHNNHAKYTYGAFVCKKCPNHFV-VDSSSCVRAKCPSSKMFV-EENGKMKCPCTDICP 179  
QY 340 RVCGYLGMOYIKANSKFIGITBLE-PAGCKKIFGSLAPLPESFPDDPASNTAPLOPEQ 398  
DB 180 KACDGIIGTSL-VSAQTVDSSNDKFNCTKNGNLFVLTGIGHDPHTTAINPEKLN 238  
QY 399 VFETLEITGYLYISAMPDLPDLSVFNQLOVIRGILHNGAYSLLTQGLGISMWGLRSL 458  
DB 239 IFQTVREITGYLNIQSPENMTDFRVSLVLTIGRSLVSGSLILKQGGITSLQFQSL 298  
QY 459 REIGSGALIHNTHLCFVHTYVPMQDLFRNPHQALLHTANREDECVGEGLAGHQLCARG 518  
DB 299 KQISAGNIYITDNSNLCYHNTVMTSLFSTPSQKTIVIRNKKAENCTADGVVNCBLSSD 358  
QY 519 HCVGPPTQCVNCSQPLRGQCEVEECRVQGLPREVYNAHCLPCHPEQCP-ONQSVTCF 577  
DB 359 GCMGPBPDOCLCKRIRIRRTICESNLYDGEFRFANQSVMECDPOCKMEDMITCY 418  
QY 578 GPRADQCAAHYKDPFVCVACPP-FNNFTVSWLVRKVASHLPEDEBACQPC 632  
DB 419 GGPBDCTCFHFKDPPNCEKPDGLQANSFIFKY-----ADEDECHPC 465  
QY 633 PINTCHSCVDLDDKGC-----PAEQASPLTISVAVV-GILLVVLGVVFGIL 680  
DB 466 HPRCTGCGRPAHSDCIYPTWRQSTLPQAR-TPL-IAAGVIGGLFIVVIGLTFAY 522  
QY 681 IKRQOKIRKYTRMLLOETELVEPLTPSGAMPNQAOMELKTELRYKVLGSAFGTV 740  
DB 523 VRRKSIR-KKRALRRL-ETELVEPLTPSGTAPNQALITKELTKRYKVLGSAFGTV 580  
QY 741 YGCIWITPDGENYKIPAIKYLRENTSPKANKELDEBAYVAGVSGYVRLGICLTSV 800  
DB 581 YGCIWVBEETVYKIPAIKILNETTGPKANVEFMDALIMASMDHPLVRLGICLSPI 640  
QY 801 QLVTLMPYGLLDHVENRGRGLGSDLLNMCQIAKMSYLEVALVHRDLAARNVLYK 860  
DB 641 QLVTLMPHGCULDYHKKNDNGSOLLNMCQIAKMSYLEVALVHRDLAARNVLYK 700  
QY 861 SPNHVKITDFGLARLLDETEYHADGKVPIKMALESILRRFTHSGDWSYGVYTW 920  
DB 701 SPNHVKITDFGLARLLDETEYHADGKVPIKMALESILRRFTHSGDWSYGVYTW 760  
QY 921 LMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMYKCMWIDECAPREFLYS 980  
DB 761 LMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIMYKCMWIDECAPREFLYS 820  
QY 981 EFSRMARDPQRFVYIONED-LGPASPLDSTFRSLJEDDDMGDLVDAEYVLPQGFPC 1039  
DB 821 EFSRMARDPQRFVYIONED-LGPASPLDSTFRSLJEDDDMGDLVDAEYVLPQGFPC 879  
QY 1040 DPAFGAGVWVHHRSSSTRSGGDLTLG-EPSEEEAPRS--PLAP-SEGASDVVDGDL 1096  
DB 880 PPIYSTRIRIDSRNRFVYRDGYAAEQV-PMRYTAPGCIIPBAFVAAGATAELFEDTC 938  
QY 1097 GNGAAGLQSLPTHDSPLQRYSEDTVP-PS-----ETDGYVAPLTCSPQPEYVNP 1149  
DB 939 CNGTLKQVATLAKEDSSTQRYADPTVFIPEVINGLEDDEGYTMPMDKXTDYLVN 998  
QY 1150 DVNPPQPSRREGFLPAP-PPAGATLERATLSGKGVKVDVDF-----AFGAVENP 1200

DB 999 EENPFVRRKXKQDQAVDNPEYRN-----ABRGQPKADEYVNEPLVNTFAETLENA 1051  
QY 1201 EYLTPOGGAAPQHPHPPAFSPARDNLVYMDQDPEEGA--PSTFPGTPI----- 1248  
DB 1052 EYL-----KNLPEKAKKAFDNPYMHSLPSTLOHPYLOESTYKPYKONGRI 1103  
QY 1249 -----AENPEYL 1255  
DB 1104 RPIVAENPEYL 1114

RESULT 6  
P79754  
ID P79754 PRELIMINARY; PRT; 1328 AA.  
AC P79754;  
DT 01-MAY-1997 (Tremblrel. 03, Created)  
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Exdb3.  
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
OC Tetraodontidae; Takifugu.  
OX NCBI\_TaxID=31033;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=99177347; PubMed=10077531;  
RA Gellner K., Brenner S.;  
RT "Analysis of 148 kb of genomic DNA around the wrt1 locus of Fugu  
rubripes";  
RL Genome Res. 9:251-258(1999).  
DR EMBL; AF056116; AAC34391.1; -.  
DR HSSP; P11362; IFG.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_Dkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_kinase.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; Dkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR ProDom; PD000001; Euk\_Dkinase; 1.  
DR SMART; SM00261; Fu\_3.  
DR SMART; SM00219; TykC; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
KW ATP-binding; Transferase.  
SQ  
SEQUENCE 1328 AA; 148613 MW; A333039258B647E9 CRC64;

Query Match 33.0%; Score 2259; DB 13; Length 1328;  
Best Local Similarity 39.7%; Pred. No. 7.7e-162;  
Matches 514; Conservative 157; Mismatches 422; Indels 202; Gaps 33;

QY 9 WGLLIALLP--GAATQ---VCTGDMKLRTPASPEHLDMRLHYQCCQVYQNLLEL 62  
DB 4 WRLIMCVASLRASSAQTEAVCPQNGLSSTSGSENGYNLXKDKYKCEIIMGLEI 63  
QY 63 TYLTPTNASLSFLQDIQVQGYVLIAHQVQVPLQRLIRYRGTLQFEDNTALAVLDNDP 122  
DB 64 TQIESNMDPFLTKIRIVETGVLLIMNHFOEIPLGQLRVIRGNSLYERRFALSFTLN--- 120  
QY 123 LNNTPPTGASPGGLREIQLRSLTEILKGGVLQGRNQQLCYQDTILMKDIFHKNNQALAT 182  
DB 121 ----YPKDG--PSGLNQLGLMNLTEIILDGVOIINKYLYGGPVYTRDII-RNNDAPIE 173  
QY 183 LIDTNRSAHCPGSPMKGSRGWSESSDQSLTRIVCAGC-ARCGPLPTDCCHEQA 241  
DB 174 IQFNGERGVCV---KSC-GNYCWPGRKQCOILTKIVCAQCDRCFGTSPRDCHECA 229  
QY 242 AGCGPGRHSQCLALHNSGICELHCPALVTYNTDFFESMPREGYTFGASCVTAAPY 301  
DB 230 AGCGPLDPTDCAFRLFNDSGACVPCQTLINNKQTFQMETPNNAKYQYGSICVSQCP 289

QY 302 NYLSTDVSGCTLVCPHNOEYTAEDGTOR-CEKSGKPCARVCYGLGMOYITANSKFIGIT 360  
 Db 290 HFV-VDGSSCVSVCPDMEV--EKGSGQCELSGCLPKYCEGTGAE---QROVDSS 342  
 QY 361 ELE-FAGCKKIFGSLAFPESEFDDPASNAPLOPEQLQVETLEETIGYLYISAMPDSL 419  
 Db 343 NIDSEINCKTQSGSHFLVLTGLDDFKNVPPPLAKKLEVRTRYREITDILNISKREL 402  
 QY 420 PDLVFNOLQVIRGRILHNGAVSLTQGLISWGLRSRLSGGLALIHNTLCEVHT 479  
 Db 403 NDLVSFSLTTIQGSLFKRFSLMWMRIPTLTSLRSLREISGSVYISQNALCYHHT 462  
 QY 480 VPMOQLFNNPH-QALHTANPEDCEVGEGLAQOLCARGCMWGPFGPCNCGOFLRGQ 538  
 Db 463 VNMTOLEFGRVRAKSLNSNRPMAECVADGRVCPDSCSGCMGWPQCCSLKNYSHG 522  
 QY 539 ECVEECVYLOGLPREYVNAH-CLPCHCECPONGSVTCFGEPAEDQVACAAYKDPFCV 597  
 Db 523 TCVAGCHFNISGIPREFAGLNGVCVACHPECKPQTKASCTGPGADECACTKPRDGPYCM 582  
 QY 598 ARCP--FNNFTVSFWLRVPKVSASHLEPDEGAQOCPINCTHSCVLDLDDKGCABEQAS 655  
 Db 583 SSCPAGVNDGEXGLIFKE-----PNREGHEPCQONCTQSGSGGLNDG--LEAA 630  
 QY 656 PLTSIVSAVVGILLVVLGVVF-----GILIKRQCKIRKTYMRRLQETELVEPLTP 708  
 Db 631 RLTISSGQITGALGVPAGLIFCLVLPFLGMLYHRLGALIRKRAMRLESGESFEPLGP 690  
 QY 709 SGAMPNQAOMRLKTEBELRKVKYVLSGAFGVYVYIGIMPDGENVKIPVALIVLENTSPK 768  
 Db 691 -GEGKTGHARILKPSDRKIKPLSGSVGFVGSXGFWIPBEETVYKIPVALITDSSGRQ 749  
 QY 769 ANKEILDEAVYVWAGSPYVSRLLGICLSTVOVLTOMPGVCLLDHREHRLGSLQDL 828  
 Db 750 TPTETDULSMGSLDHPYIVRLGICGTCLOLVLTQSSGSLSEHROKTKISLDPQL 809  
 QY 829 LNMCMQIAKMSYZEDVLVHRDLAARNVLKSPNHVKITDFGLARLIDDETYHADGG 888  
 Db 810 LNMVQIAKMWYLEEHVYHKNLAARNILKNDYQVLSQYVADLLYPDDKKYVYSET 869  
 QY 889 KVPITKMALESILRRFTHQSDVWSYGVTVWELTFGAKPFDGIPAREIPLEKGERLP 948  
 Db 870 KPIKMMALLESILFRYTHQSDVWSYGVTVWEMMSFGAEPYASVQPEVPSVLEKGERLS 929  
 QY 949 QPCTIDIVYIMVCMWIDSECRPRELSEFSRMAKDPQRFVVIQNEIDLPASPLDS 1008  
 Db 930 QPACTIDIVYVWVCMWIDENIRPTEKELASDTFNARADPRYLIVIMEG----- 980  
 QY 1009 TFRSLLEDDMDGLVDAEYLVPOQGFCCPDPAAGAGVHHRSSSTRSGGDLTLG 1068  
 Db 981 -----EDSGMGERL-----RGSER--GLEAD 1001  
 QY 1069 LEPEEAPRSPPLAPSEGAGSDVFDGLGMG--AAKGLOSLPTHDSPLD----- 1116  
 Db 1002 LEEDEE-----GLSDRAFTSLQPSWSWSPSINSYVMTQ 1040  
 QY 1117 -RYSEDPTVPLSETDGYVALTCSPOP-EVYNQ-----PVYRPOPSPRE 1160  
 Db 1041 LRYD-----FAVQGGHIGYLPMSPSBVDITRQMTQSRSLSVTLPRDSARRSRE 1094  
 QY 1161 GPL--PAAPAGATLERAKTILSPKNGVYKVFAGAVENPEVLTQGGAAAPPHPPA 1218  
 Db 1095 AELCEDGAQACGIRVR-----FGSERG--POGG----- 1122  
 QY 1219 FSPAFDNLVYWDQDPPERGAPSTFGTPTAENPE 1253  
 Db 1123 -----QORKLSTASSPSSEFKTMADEDE 1146

RESULT 7  
 Q9BIH9 PRELIMINARY; PRT; 1433 AA.  
 AC Q9BIH9;

DT 01-JUN-2001 (TEMBUREL 17, Created)  
 DT 01-JUN-2001 (TEMBUREL 17, Last sequence update)  
 DT 01-JUN-2002 (TEMBUREL 21, Last annotation update)  
 DE Putative epidermal growth factor receptor (Fragment).  
 GN EGR.  
 OS Anopheles gambiae (African malaria mosquito).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;  
 OC Anopheles.  
 NC NCB1\_Taxid=7165;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SLA;  
 RA Lycett G.J.;  
 RT "Cloning, expression and localisation of the Anopheles gambiae  
 RL epidermal growth factor receptor."  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ301655; CAC35008.1;  
 DR HSSP; P13362; 1FGX  
 DR InterPro; IPR000345; Cyt\_c\_heme\_bind.  
 DR InterPro; IPR000494; EGR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00757; Furin-like.1.  
 DR Pfam; PF00069; kinase.1.  
 DR Pfam; PF01030; Recep\_L\_domain.2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_kinase.1.  
 DR SMART; SM00261; Fuf\_7.  
 DR SMART; SM00220; S\_TKc.1.  
 DR SMART; SM00219; TYKc.1.  
 DR PROSITE; PS00180; CYTOCHROME\_C; UNKNOWN\_4.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_AFP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW Receptor.  
 FT NON TER.  
 SQ SEQUENCE 1433 AA; 159585 MW; 83D9D88967724F07 CRC64;  
 Query Match 29.6%; Score 2026.5; DB 5; Length 1433;  
 Best Local Similarity 32.8%; Pred. No. 3.5e-144;  
 Matches 474; Conservative 199; Mismatches 395; Indels 379; Gaps 39;  
 QY 26 CTCTDKKRLPASPETHLMLRLYGCQVQGNLELTYLPTNASISFLQIDQVGGYVL 85  
 Db 1 CIGTNGMSVPAAREHYKRLRDRYNTCTVYGNLEITWIONITDINFLQHIRVYGYVL 60  
 QY 86 IAHNOVROYPLQRLRIVRGOLF-----EDNYALAVLDNGDPLNNTPTVGASPGGLREL 140  
 Db 61 ISLYDLPOVILPRLQIIRGRITTKLNKMEAGLVF-----SFSMNTL 104  
 QY 141 QLSRLTEILKGVLIQARNPOLCYQDTILMKDI-FHKNOGLATLIDTNSRACHPCSPMC 199  
 Db 105 ELPALRDLIGSGVGFNNYNLCHMKXINNEIILAPQTSQMTFNSSPERVCPCHPSC 164  
 QY 200 KKSRCMGSESDOSLITRVACGCA--RCKRPLPDDCHCQCAAGCTPRKSDICACH 257  
 Db 165 EVC-CMGEQGHNCQRFKSLNCSFQSGCGCFGRKPECHLFCAGGCTGTPQSDCLACN 223  
 QY 258 FHNISGICELCPALVTYNNDTFESMNPREGRYTFGASCTACAPYNYLSTDVSGCTLVCP 317  
 Db 224 FYDDGVCKQECPEPMQYNPTVFWENPBGKAYGATVCRKCP-EHLLDNAGACVAKCKX 282  
 QY 318 HNOEYTAEDGTORCEKCSFPCARVCYGLGMOYITANSKRTGTELEFACCKIFGSLAF 377  
 Db 283 GKMPONS-----CVPCKGVCPKTCPEGBI-----VHSDNIG-----NYKCTIIESLETL 329  
 QY 378 PESFDDPASNT-----APLOEPOLOVFETLEETIGYLYISAMPDSLPLDSVFNQ 429  
 Db 330 DQSGDFQGVYNNFSGPRYIKIDPRLEVFSTVKEITGFINIQAHHFPTLNTNFRNLE 389

```

QY 430 VIRGILLINNGAY-SLTQGLGISWLGRLSEGLGSLTLHHNHLGCVHTVPMWDLFRN 488
DB 390 VVGRQLKXENIIFASVYVYKTSLSKLELSKRVNNGSIVLENSDLTFVEDIDMSLKKS 449
QY 489 PHQALHTANRPEDECEVGEGLACHOLCARGHCMWPGPTQCVCNCSQPLRGCEVECEVLQ 548
DB 450 SDHEVMVQKNNATECHEGEGECGSCSKAGCMKGPBQCLCNVYKXKGLDSCX--- 506
QY 549 GLPREY-VNANHCLPCHPECCOPONGSVTCFEPADQCVACHYHDPFCVACRCP----- 601
DB 507 SLFRLYSVDSKTCDDCHQOECD-----FCYGPBNEDNGSCNNVXDRFRCAECPTTHAM 561
QY 602 ---FNNFVFFMLRVK-----VSASHLE-----PD----- 624
DB 562 NGTCNCHKTCCVGRPRDTIAPDGCISCDAKIIIGSDAKIERCLMDSCDGDYSDYVL 621
QY 625 -EEG----- 627
DB 622 QEEGFLKQISGKAVCRKCHPRCKCTGYGFHEGFCOECTGYKGECEDECPDPFVANE 681
QY 628 -ACOPCPINCT-----HSCVDL-----DP-----KGPBAEQ----- 652
DB 682 TRICLPHQOEGRGHGLGDHHECRNLKLFEGDGYDNATTTTCVSNCPASHRYRFPQEA 741
QY 653 -----RASPLTSVSAVVGILLVVLGVVFGI---LIKROQKIRRYTM 693
DB 742 GKIGPYCSADSMOSGLRIEPOQVYKIVWGSVMALLLCVFGIAGFVLFSSHKNKDAVKM 801
QY 654 RLLQETELVPLTPSGAMPNOAMRIKETELRKVYLSGARGATYKGIWIPDGNVK 753
DB 802 TMLAGCDSSEPLRPSNVGPNLTKRILKEAIEIRGGVLMGAGRGFRKGYMNEGSESVK 861
QY 754 IPAIKVLRNTPSPKANKEILDEAYVNAVGSFVYSRLGICLTSTVOLVTOLMPYGLL 813
DB 862 IPAIKVLMENSGSSSKSEPLEEAYIMASVEHPNMLKLLAVCMTSQMMLLITQMLPGLCL 921
QY 814 DHVRENRRRLSGQDLNMCQIAGKMSYLEDVRLVHDLARNLVYSPNHVKTIDEGLA 873
DB 922 DYVRNKKDKISSKALLNWSQIARGMALIERRLVHDLARNLVYTPSCVKTIVTGLA 981
QY 874 RLIDIDETEHADGGKXVPIKMMALLESILRRFTQSDVWSYGVTVWELMTEGAKPYGIP 933
DB 982 KLDPFDSDEYRAGSKMPTKMLALECIHRVFTSKSDVMARGLITIMELTIGARPYENVP 1041
QY 934 AREIPDLKEKEBERLPORPITCTIDYMTVMKCMADSECRPFRELVSEFSRMAPDPQFV 993
DB 1042 AKDYVELIEIGHKLPQDPICSLDYVCIILSCWVLDADARPTFKQLAETFAEKADPGRYL 1101
QY 994 VIQMEDLAGASPLDSTFYRSLLEDDMDGLV----- 1024
DB 1102 MI-----FGDKFMRLLPSYTNODEKDLRTLAAYMAAALAAAGASNVDPSTIA 1152
QY 1025 DAEYLVPOQGFCCDPAPGAGMVRHHRSSSTRSGGDLTLGLBSEEBEPAS----- 1079
DB 1153 ETDEYLTQKTRPSIMLPGPSA-----VEBS-DEMPKSLRYCK 1188
QY 1080 -PLAP---SEGASDVFPDGLGMGAKGLQSLPTHDSPLGRYSEDPVPLPSTDDYVA 1135
DB 1189 DPLKPDDETDGHEV-----GVGIR-----LNLPLEDDDTL 1222
QY 1136 PLTCSPOPEYVQDPVDPQPPSPREGPLPAARFAGATLERAKTSLPGKNGVYCVAFGG 1195
DB 1223 P-TCOSQ---NQS-----TPG-----YMDLIGVPA 1243
QY 1196 AVENPEYL-----TPGGAAPQHPHPPAPSPAPFNDLYWMDPPRERCARPSTFKGT 1246
DB 1244 SVDNPEYLMGSTQAIAGLAQSGM--PHTPP-----PNTNNGM 1280
QY 1247 PTAENPE 1253
DB 1281 PTHQHSQ 1287

```

```

RESULT 8
ID Q9UK79 PRELIMINARY; PRT; 419 AA.
AC Q9UK79;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10465918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor."
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR SMART; SM00261; Fur_1_domain; 1.
SQ SEQUENCE 419 AA; 45472 MW; FECLIBE347E2D030C CR664;

Query Match 27.3%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 3.3e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 METALCRGGLLALLPRGAATQVCTGDMKRLRPASPTLDMRLHYQCCQVQVGNL 60
DB 1 METALCRGGLLALLPRGAATQVCTGDMKRLRPASPTLDMRLHYQCCQVQVGNL 60
QY 61 ELTYLPTNLSLFLDIOGVQGVLLAHNQVQVPRQRLIRVGTQLFEDNYALAVLNG 120
DB 61 ELTYLPTNLSLFLDIOGVQGVLLAHNQVQVPRQRLIRVGTQLFEDNYALAVLNG 120
QY 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNOLA 180
DB 121 DPLNNTPTVGTASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNOLA 180
QY 181 LTIIDNRRBACHPGSPMKSGRCWSESESDOSLRTYACAGGACRCKGPLPTDCHEOC 240
DB 181 LTIIDNRRBACHPGSPMKSGRCWSESESDOSLRTYACAGGACRCKGPLPTDCHEOC 240
QY 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTRESMPNREGRTYFGASCYTACP 300
DB 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTRESMPNREGRTYFGASCYTACP 300
QY 301 YNLTSDVGSCTLVGCIHNOEYTAEPGTGRCKSGSPCARVYCYGL 345
DB 301 YNLTSDVGSCTLVGCIHNOEYTAEPGTGRCKSGSPCARVYCYGL 345

RESULT 9
ID Q8R2X1 PRELIMINARY; PRT; 367 AA.
AC Q8R2X1;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;

```

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC027080; AA027080.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;  
 Query Match 25.4%; Score 1739; DB 11; Length 367;  
 Best Local Similarity 88.0%; Pred. No. 2,6e-123;  
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;  
 QY 895 MALESILRRRFTHSGDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 954  
 DB 1 MALESILRRRFTHSGDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60  
 QY 955 IDVYIMVWCMMIDSECRPRRELVSERSMARDQRFVIONEDLGASPLDSTFYRSL 1014  
 DB 61 IDVYIMVWCMMIDSECRPRRELVSERSMARDQRFVIONEDLGASPLDSTFYRSL 120  
 QY 1015 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 1074  
 DB 121 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 180  
 QY 1075 EAPRSPPLASSEGAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPTVLPSETDGYV 1134  
 DB 181 EAPRSPPLASSEGAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPTVLPSETDGYV 240  
 QY 1135 APLTCSPOPEYVNOVDVPRQPSPREGFLPARPAGATLERAKTILSPGKGVKDVFAFG 1194  
 DB 241 APLTCSPOPEYVNOVDVPRQPSPREGFLPARPAGATLERAKTILSPGKGVKDVFAFG 300  
 QY 1195 GAVENPEYLTPOGGAAPQPPAPSPAFDNLVYWDPPERGAPSTFGPTTANPEV 1254  
 DB 301 GAVENPEYLTPOGGAAPQPPAPSPAFDNLVYWDPPERGAPSTFGPTTANPEV 360  
 QY 1255 LGLDVVPY 1261  
 DB 361 LGLDVVPY 367  
 RESULT 10  
 Q8WYVO  
 ID Q8WYVO PRELIMINARY; PRT; 412 AA.  
 AC Q8WYVO;  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical 44.7 KDa protein.  
 GN PP3659.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,  
 RA Wan D.F., Gu J.R.;  
 RT "Novel human cDNA clones with function of inhibiting cancer cell  
 RT growth."  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF18349; AAL55856.1; -  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00219; Tyrc; 1.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Hypothetical protein.

SQ SEQUENCE 412 AA; 44702 MW; 034397FF327D2BC CRC64;  
 Query Match 24.8%; Score 1697.5; DB 4; Length 412;  
 Best Local Similarity 80.5%; Pred. No. 4,4e-120;  
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;  
 QY 895 MALESILRRRFTHSGDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 954  
 DB 1 MALESILRRRFTHSGDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICT 60  
 QY 955 IDVYIMVWCMMIDSECRPRRELVSERSMARDQRFVIONEDLGASPLDSTFYRSL 1014  
 DB 61 IDVYIMVWCMMIDSECRPRRELVSERSMARDQRFVIONEDLGASPLDSTFYRSL 120  
 QY 1015 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 1074  
 DB 121 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 180  
 QY 1075 EAPRSPPLASSEGAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPTVLPSETDGYV 1134  
 DB 181 EAPRSPPLASSEGAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPTVLPSETDGYV 240  
 QY 1135 APLTCSPOPEYVNOVDVPRQPSPREGFLPARPAGATLERAKTILSPGKGVKDVFAFG 1194  
 DB 241 APLTCSPOPEYVNOVDVPRQPSPREGFLPARPAGATLERAKTILSPGKGVKDVFAFG 300  
 QY 1195 GAVENPEYLTPOGGAAPQPPAPSPAFDNLVYWDPPERGAPSTFGPTTANPEV 1254  
 DB 301 GAVENPEYLTPOGGAAPQPPAPSPAFDNLVYWDPPERGAPSTFGPTTANPEV 360  
 QY 1227 YWMD-QDPER-----GAPSTFGPTTEN 1251  
 DB 361 WWTQCEBEGQVRSPDVSSGREGLTISAGIKMEGCPPTISRGTHARN 410  
 RESULT 11  
 Q86712  
 ID Q86712 PRELIMINARY; PRT; 729 AA.  
 AC Q86712;  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 DE Polypeptide.  
 GN POLYPROTEIN.  
 OS Avian rous-associated virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.  
 OC NCBI\_TaxID=11950;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94203659; PubMed=8152791;  
 RA Vennstrom B., Raynocheck C., Jansson L., Doederlein G., Lhotak V.,  
 RA Johnson A., Beug H.;  
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
 RT different transforming capacities."  
 RL Oncogene 9:1307-1320(1994).  
 DR EMBL: S69372; AAC60725.1; -  
 DR HSP: P03322; 1A6S  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR004028; Retro\_M.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF02813; Retro\_M; 1.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00219; Tyrc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Tyrosine-kinase; Tyrosine-protein kinase.  
 SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EFD1D63 CRC64;  
 Query Match 24.5%; Score 1679; DB 15; Length 729;  
 Best Local Similarity 53.8%; Pred. No. 2,6e-118;

Matches 357; Conservative 71; Mismatches 119; Indels 116; Gaps 17;

QY 569 PONSATCGEADQCAAHYKDPFCVACRCP---NNFTVSFWLRVPRKVSASHLEPD 624  
 DB 141 PEETATPKTGP--DHCKCAHAFIDGPHCVACRCPGVLGENDTL-VWKYA-----D 187  
 QY 625 EEGACQCPINCTHSCVDLDKGCAPQARASPLTISYSAV-GILLVVLGVFGILIKR 683  
 DB 188 ANAVCOLCHPNCIRGCGKPGLEGCP---NGSKTPSIAGVGGILCLVVGILGLYLRR 244  
 QY 684 RQCKIRKRYMRRLIQETELVEPLTPSGAMPQAOMLIKETELRKVKVLSGAFYVYK 743  
 DB 245 R-HIVRKRTLRLQERELVEPLTPSGEAPQOALRIKETEFEKVKVLSGAFYVYK 303  
 QY 744 IMIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVWAGSPVSRLLGICLTSTQV 803  
 DB 304 LMPEGEKVKIPVAIKELRENTSPKANKELIDEAYVWASVNDPVCCLLIGLITVQ 363  
 QY 804 TQLMPYGCLLDVHVENRGRLSODLLNMCQIAKMSYLEVDVLRDLAARVLYKSPN 863  
 DB 364 TQLMPYGCLLDYIREHKDNIGSYLLNMCVQIAKMSYLEERLVRDLAARVLYKTPQ 423  
 QY 864 HVKITDPCGLARLLIDETEHYADGKVPKIMMALESILRRRFTHQSVDVSGVTVWELMT 923  
 DB 424 HVKITDPCGLARLLIDETEHYADGKVPKIMMALESILRRRFTHQSVDVSGVTVWELMT 483  
 QY 924 FGAKPYDGIAPAREIPDLLEKGERLPQPPICITIDVYIMVWKMMIDSECRPFRELYSEFS 983  
 DB 484 FGSKPYDGIAPAREIPDLLEKGERLPQPPICITIDVYIMVWKMMIDSECRPFRELYSEFS 543  
 QY 984 RMAADPQRFVYIQ-NEDLGPRASPLDSTFYRSLEDDMDGLVDAEYLVPQCGFCDDPA 1042  
 DB 544 RMAADPQRFVYIQ-NEDLGPRASPLDSTFYRSLEDDMDGLVDAEYLVPQCGFCDDPA 598  
 QY 1043 PGAGMVAHHRHSSTSRSGGDLTLGLEPSEEARPSPL-----APSEGAGDVFDGDLG 1097  
 DB 599 -----NSPST-----SRTPLLSLSATSNNAATNCID----- 625  
 QY 1098 MGAAGLQSLPTHPSPLORYSEDPTVPLPSET--DGVAVALTCSPOPEYVNOQDVAPQP 1155  
 DB 626 ---RNGCGHVRSDSEFVQRYSSDPTGNFLEESIDDFL-----PAPEYVNO--LMPKK 673  
 QY 1156 PSPEGGLPARAPGATLERAKTSLSPGNGVGVKDF-----AFGAVENP 1200  
 DB 674 PS-----TAMVQNIYNNISLTALSKLPMDSRYQNSHSTAVDNP 712  
 QY 1201 EYL 1203  
 DB 713 EYL 715

RESULT 12

Q66714 PRELIMINARY; PRT; 567 AA.  
 AC Q66714;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE V-erbB protein (Fragment).  
 GN V-ERBB.  
 OS Avian rous associated virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.  
 OX NCBI\_TaxID=11950;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RX MEDLINE=94203659; PubMed=8152791.  
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
 RT different transforming capacities."  
 RL Oncogene 9:1307-1320(1994).  
 DR EMBL; S69372; AAC60727.1; -.

DR HSP; P11362; 1FGK.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00219; TYRK; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW Tyrosine-protein kinase.  
 FT NON\_TER  
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7AD725E1 CRC64;

Query Match 24.5%; Score 1677; DB 15; Length 567;  
 Best Local Similarity 54.4%; Pred. No. 2.5e-118;  
 Matches 356; Conservative 69; Mismatches 113; Indels 116; Gaps 17;

QY 578 GEADQCAVCAHYKDPFCVACRCP---NNFTVSFWLRVPRKVSASHLEPDDEGACQCP 633  
 DB 1 GP--DHCKCAHAFIDGPHCVACRCPAGVLGENDTL-VWKYA-----DANAVCOLCH 47  
 QY 634 INCTHSCVDLDKGCAPQARASPLTISYSAV-GILLVVLGVFGILIKRQCKIRKT 692  
 DB 48 PNCITGCKPGLEGCP---NGSKTPSIAGVGGILCLVVGILGLYLRR-HIVRKRT 103  
 QY 693 MRLLQETELVEPLTPSGAMPQAOMLIKETELRKVKVLSGAFYVYKGIPIPDGENV 752  
 DB 104 LRRLLQERLVEPLTPSGAMPQOALRIKETEFEKVKVLSGAFYVYKGIPIPDGENV 163  
 QY 753 KIPVAIKVLRNTSPKANKELIDEAYVWAGSPVSRLLGICLTSTQVLTQLMPYGCL 812  
 DB 164 KIPVAIKELRENTSPKANKELIDEAYVWASVNDPVCCLLIGLITVQLMPYGCL 223  
 QY 813 LHVVENRGRLSODLLNMCQIAKMSYLEVDVLRDLAARVLYKSPNKTITDGL 872  
 DB 224 LDIYRHKDNIGSYLLNMCVQIAKMSYLEERLVRDLAARVLYKTPQHVKITDPCGL 283  
 QY 873 ARLLIDETEHYADGKVPKIMMALESILRRRFTHQSVDVSGVTVWELMTFGAKPYDGI 932  
 DB 284 AKLLGADKEHYAEGKVPKIMMALESILRRRFTHQSVDVSGVTVWELMTGSKPYDGI 343  
 QY 933 PAREIPDLLEKGERLPQPPICITIDVYIMVWKMMIDSECRPFRELYSEFSMADDPQF 992  
 DB 344 PASELSSVLEKGERLPQPPICITIDVYIMVWKMMIDSECRPFRELYSEFSMADDPQF 403  
 QY 993 VVIQ-NEDLGPRASPLDSTFYRSLEDDMDGLVDAEYLVPQCGFCDDPAAGAMVH 1051  
 DB 404 LVIQGERHRLSPDTSKRYRTLMEEDMEDIVDAEYLVPQCGF----- 449  
 QY 1052 RHRSSSTRSGGDLTLGLEPSEEARPSPL-----APSEGAGDVFDGDLGGAAGLQ 1106  
 DB 450 --NSPST-----SRTPLLSLSATSNNAATNCID-----RNGG 481  
 QY 1107 LPTHPSPLORYSEDPTVPLPSET--DGVAVALTCSPOPEYVNOQDVAPQPSPREGPLP 1164  
 DB 482 HVRSDSEFVQRYSSDPTGNFLEESIDDFL-----PAPEYVNO--LMPKKS----- 526  
 QY 1165 AARPAATLERAKTSLSPGNGVGVKDF-----AFGAVENPEYL 1203  
 DB 527 -----TAMVQNIYNNISLTALSKLPMDSRYQNSHSTAVDNPYL 566

RESULT 13

Q64895 PRELIMINARY; PRT; 962 AA.  
 AC Q64895;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Gag v-erbA, v-erbB protein.  
 GN GAG-V-ERB-A, V-ERB-B.  
 OS Avian erythroblastosis virus.



OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.  
 NCBI\_TaxID=11861;  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=90206603; PubMed=1969616;  
 RA Brinkman A., Jackson J., Bishop J.M., McCarty D.J., Schltzman R.C.;  
 RT "Six amino acids from the retroviral gene gag greatly enhance the  
 transforming potential of the oncogene v-etb-B";  
 RL Oncogene 5:15-24(1990).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 DR EMBL; X52209; CA36459.1; -  
 DR EMBL; X52211; CA36459.1; JOINED.  
 DR HSSP; P10828; NTL.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR000536; Hormone\_rec\_1ig.  
 DR InterPro; IPR001723; Stdhm\_recptor.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR InterPro; IPR001628; Znf\_Cstereoid.  
 DR Pfam; PF00104; hormone\_rec; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00105; Zf-C4; 1.  
 DR PRINTS; PR00398; STRDHOMONER.  
 DR PRINTS; PR00047; STROIDFINGER.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR ProDom; PD000035; Znf\_Cstereoid; 1.  
 DR SMART; SM00430; HOL1; 1.  
 DR SMART; SM00239; TykC; 1.  
 DR SMART; SM00399; Znf\_C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; DNA-binding; Nuclear protein; Receptor;  
 KW transcription regulation; Transferase; Tyrosine-protein kinase;  
 KW Zinc-finger.  
 SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E495CE CRC64;  
 Query Match 23.6%; Score 1612.5; DB 15; Length 962;  
 Best Local Similarity 50.9%; Pred. No. 4.2e-113;  
 Matches 357; Conservative 69; Mismatches 143; Indels 133; Gaps 20;  
 QY 541 VEECRVLQGLPRE-YVNAR-HCLP-----CHPECC 568  
 DB 1EKQGESYLLAFEHYINVRKHNIHFWSKLMKVADLMIGAYHASFIMKXCEPTELS 413  
 QY 569 PONGSVTCFGEADQCVCAHYKDPFCVACPF---NFTVSFWLRVKVSAASHLEPD 624  
 DB 414 PQE-----VGP--DHCKCAHFIDGPHCVACPAVGIGENDTL-WMKYA-----D 455  
 QY 625 EEGACQCPINCTHSCVDLDKGCFAQRASPLTSIVSAVY-GILLVVLGVFGILIKR 683  
 DB 456 ANAVCQCHPCNCTRGCGPGLEGCP---NSKTPSIAAGVVGILCLVVGIGLYLRR 512  
 QY 684 RQCKIRKVTMRLLQETELVEPLTPSGAMPNOAOMRIKTELKRVKVLGSGAFVYK 743  
 DB 513 R-HIVKRTLRRLQERELVEPLTPSGEAPNOAHLRIKTEFEKRVVLGFGAFVYK 571  
 QY 744 IWPIDGENVKIPVAIKVIRENTSPKANKELIDEAYVAGVSPVYSRLGICLTSTVOLV 803  
 DB 572 IWPIDGENVKIPVAIKVIRENTSPKANKELIDEAYVAGVSPVYSRLGICLTSTVOLV 803  
 QY 804 TQMPYGCCLLDHVENRGRLGSGODLNMWQIAKMSYLEDVRLVHDLAARVLYKSP 863  
 DB 632 TQMPYGCCLLDHVENRGRLGSGODLNMWQIAKMSYLEDVRLVHDLAARVLYKSP 863  
 QY 864 HVKLTDFGLARLLDIDETEHADGGKVPKMMALSTLRRTFHOSDVMSYGVTVWELMT 923  
 DB 692 HVKLTDFGLARLLDIDETEHADGGKVPKMMALSTLRRTFHOSDVMSYGVTVWELMT 923  
 QY 924 FGAKPYDGIAPAREPDLLEKGERLPDPICITIDVYMIWYKMMIDSECRPRFRELSEFS 983  
 DB 752 FGAKPYDGIAPAREPDLLEKGERLPDPICITIDVYMIWYKMMIDSECRPRFRELSEFS 983

QY 984 RMAPDQRFVVIC-NEDLGASPLDSTFYRSLEDDMDGLVDAEELVLPQGFPCPDPA 1042  
 DB 812 KVARDPRLVIGDGRMKLPSPTDSKFRITLMEEDMDIDVADADYLVPHQGF----- 866  
 QY 1043 PGAGVNHHRHRSSTRSGGDTLLGLEPSEEBAPSPPLAPBAGSDVFDGDLGMAK 1102  
 DB 867 -----NSPST-----SRPILLSISATSN-----NSATK 890  
 QY 1103 GLQSLPTHPDPSLPORVSEDPVPLPSETDGVYAPLTCSPQEPYVNOPDVPPPPREGE 1162  
 DB 891 CIDRNGH-----PREHGFL-----PAPEVNO-LMPKPFSTAMQ 926  
 QY 1163 LPARAPAGAT-IERAKTLSPKNGVYKDVFAFAGAVENPEYL 1203  
 DB 927 NQIYNYISLTALSKLPMDSRYQN-----SHSTAVDNPEYL 961  
 RESULT 14  
 Q85468 ID Q85468 PRELIMINARY; PRT; 545 AA.  
 AC Q85468;  
 DT 01-NOV-1996 (TREMELREL. 01, Created)  
 DT 01-NOV-1996 (TREMELREL. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMELREL. 20, Last annotation update)  
 DE Avian erythroblastosis virus (Ts34) v-etbB gene.  
 OS Avian erythroblastosis virus.  
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.  
 NCBI\_TaxID=11861;  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=88217326; PubMed=2897102;  
 RA Scotting P., Vennistrom B., Jansen M., Graf T., Beug H., Haymann M.J.;  
 RT "Common site of mutation in the etbB gene of avian erythroblastosis  
 virus mutants that are temperature sensitive for transformation";  
 RL Oncogene Res. 1:265-278(1987).  
 DR EMBL; X06943; CA30024.1; -  
 DR HSSP; P11362; IREG.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00219; TykC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Transferase; Tyrosine-protein kinase.  
 KW Sequence 545 AA; 60899 MW; 140DC8C8CA0F8AF4 CRC64;  
 Query Match 23.4%; Score 1604; DB 15; Length 545;  
 Best Local Similarity 53.9%; Pred. No. 8e-113;  
 Matches 344; Conservative 66; Mismatches 122; Indels 106; Gaps 17;  
 QY 578 GREADQCVCAHYKDPFCVACPF---NFTVSFWLRVKVSAASHLEPDDEGACQPCP 633  
 DB 1 GP--DHCKCAHFIDGPHCVACPAVGIGENDTL-WMKYA-----DANAVCQCH 47  
 QY 634 INCTHSCVDLDKGCFAQRASPLTSIVSAVY-GILLVVLGVFGILIKRPOQIKRYT 692  
 DB 48 PNCTRGCGKPGLEGCP---NSKTPSIAAGVVGILCLVVGIGLYLRR-HIVKRT 103  
 QY 693 MRLLQETELVEPLTPSGAMPNOAOMRIKTELKRVKVLGSGAFVYKGIWIDGENV 752  
 DB 104 LRLLQERELVEPLTPSGEAPNOAHLRIKTEFEKRVKVLGSGAFVYKGIWIDGENV 752  
 QY 753 KIPVAIKVIRENTSPKANKELIDEAYVAGVSPVYSRLGICLTSTVOLVTLQMPYGC 812  
 DB 164 TTPVAIKVIRENTSPKANKELIDEAYVAGVSPVYSRLGICLTSTVOLVTLQMPYGC 812  
 QY 813 LDHVENRGRLGSGODLNMWQIAKMSYLEDVRLVHDLAARVLYKSPHVKITTDGL 872  
 DB 224 LDYIREHKNIGSQYLLNMWQIAKMSYLEDVRLVHDLAARVLYKSPHVKITTDGL 872

QY 873 ARLLDDETEYHADGKVPKIMMALESILRRFTHQSDVWGYVTWELMTFGAPYDGI 932  
 DB 284 AKQIGADKEKYEABEGKVPKIMMALESILHRYTHQSDVWGYVTWELMTFGSKPYGI 343  
 QY 933 PAREIPDLLEKGERLPDPICITIDVYMTWKCWMIDSCRPREFRELVSEFSRMAPDPORF 992  
 DB 344 PASEISVLEKGERLPDPICITIDVYMTWKCWMIDSCRPREFRELVSEFSRMAPDPORF 403  
 QY 993 VVIQ-NEDGPASPLDSTFFRSLLEDMDGLVDEEYLVVQQGFCCDPAPAGAWMHH 1051  
 DB 404 LVIOQDERKHLPSFDPSKFRYRLTMEEDMEDVYDDELIVPHQGF----- 449  
 QY 1052 RHRSSSTNSGGDLTLGLPSEEAERASPL-----APSEGASGVFDGDLGMAKGLQS 1106  
 DB 450 -NSPST-----SRTPLLSLSATSNMNSATNCDIRNG----- 480  
 QY 1107 LPTHPSPLQRSEDPYVPLPSETDGYVAPLTCSPQRYVNPDPVPPSPREGPLPAA 1166  
 DB 481 -H-----PVREDGFL-----PAPRYVNO-LMPKKPSTANVQIQIY 513  
 QY 1167 RPAGAT-LEBAKTLSPGNKGVKDYAFAGAVENPEYL 1203  
 DB 514 NYISLTATSKLPMDSRYQN-----SHSTAVDNEYL 544  
 RESULT 15  
 Q9WVF5 PRELIMINARY; PRT; 655 AA.  
 AC Q9WVF5; 01-NOV-1999 (TREMBLrel. 12, Created)  
 DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).  
 DE EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_Taxid:10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=LIVER;  
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Mahle N.J.;  
 RA "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors.";  
 RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinselaar C.S., Pearsall R.S., Green P.J., Yee D., Lampand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Mahle N.J.;  
 RA "Comparative genomic sequence analysis and isolation of human and mouse alternative Egr transcripts encoding truncated receptor isoforms.";  
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=LIVER;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawada T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamaoka I., Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schmitt L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Mashio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Ljunggren S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitlaker C., Wilming L., Wyshew-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S., Hayashizaki Y.;  
 RA "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 DR EMBL: AF124513; AAD44149.1; -  
 DR EMBL: AF275366; AAG28047.1; JOINED.  
 DR EMBL: AF275366; AAG28047.1; JOINED.  
 DR EMBL: AF275365; AAG28047.1; JOINED.  
 DR EMBL: AK004944; BAB23688.1; -  
 DR EMBL: AK004883; BAB23641.1; -  
 DR EMBL: AK004911; BAB23662.1; -  
 DR MGI: MGI:95294; Egrf.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR002174; Furin-like.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR SMART: SM00261; Fu; 3.  
 KW Receptor.  
 SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;  
 Query Match 21.5%; Score 1468.5; DB 11; Length 655;  
 Best Local Similarity 43.9%; Pred. No. 1,9e-102;  
 Matches 286; Conservative 90; Mismatches 234; Indels 41; Gaps 11;  
 QY 11 LIALALPPGAA--STOVCSTDMKRLPASPEHLMRLHYGGCCOVGNLELTLPNTN 68  
 DB 14 LITLALCAAGALBEKRYCQSTSNRLTQLGFEDHFLSTLRMYNCCVYVGNLEITYVORN 73  
 QY 69 ASLSPLODIOGVGYLTAHNOYRQVPLRLIVRGNOIFEDYVYALVNDGPLNNTTP 128  
 DB 74 YDSLFLTKTIOEVAGVYLIALNTYERPLENTQLIRGNALYENTYALAILSN----- 124  
 QY 129 VTGASPGLEIRELQSLSTTEILKGVLIQRNPOLCYODTILMKDI---FKKNQLALTL 184  
 DB 125 -YGTNRFLGRELPMRNLOETLIGAVFNSNPILCNMDTQMRDLYGVNFVSNMSMDL--- 180  
 QY 185 DTRRSACRCHSCMCKGSRGWSSSPDCQSLRTVAGGCA-RCKRPLPTDCHDCCAG 243  
 DB 181 -QSHPSCKPCDSCPNWSCGSGEENCCQLTKITICAQCSHRCRSPSDCCNQCAG 239  
 QY 244 CTGPKHSDCIACLFHNSGICELHCPALVTYNTDPESMNPGRATFGASCYTAAPYNY 303  
 DB 240 CTGPRSDCLVCKPFQDEATCKDTCRPLMLYNPTTYQMDVNPBGKVSFGATCYKCPRYN 299  
 QY 304 LSTVDGSCITVCEPLHNOEYTAEDGTORCEKSPCARVCGYGLMOYIKANSKFTGLE 363  
 DB 300 VVTDHGSVCVACCPDYEV-EEGDIKKCKCDPRKVCNNGIIGFFK-DTSLINATNKK 357  
 QY 364 -FAGCKKIFGSLAFLESPDPGASNTAPLOPQOVFETLEITVLYIASPDSLPDL 422  
 DB 358 HFXYCTAISDHLILVAFKGDSPFTTPPLDRELILKTVKEITFFLLIQAMPDWTL 417  
 QY 423 SVFQNTQVTRGRIHNGAVSLTQLGIGISMLGRLSRLDGSGIALIHNTHLCEVTPW 482  
 DB 418 HAFENLEIIRGTRKQGFSLVAVGINTSLGIRLSKEISDGDVILISGRNLCYANTIM 477  
 QY 483 DQLEFRPHOALHTARPEDECEVGEGLAHO-CARHCGGPGTQCVNCSQFLRQECVE 542  
 DB 478 KKLFGPNQTKTMMNRAEKDCAVNVHVCNPLCSSGCGPBRDVCSCQNSRGEVCE 537  
 QY 543 ECRVQLGPREVYNAHCLPCHPCQPNQSVTCFQPEADQCAAHYDPFCVAPCPF 602  
 DB 538 KCVILGEBREFEVNESECICQHECPLQAMNITCTGRGPDNCIQCAHYIDGPHCVKTCGA 597  
 QY 603 -----NNFTVSPFLRKPXKASLLEPDEGACQPCPGINTHSCVDLDIDGCG 648  
 DB 598 GIMGENNTLV--MKYA-----DANNVCHLCNANTCYGAGGGLGCG 636

Tue Jul 22 12:41:42 2003

seq4-579-593-14.rpt

Page 13

Search completed: July 22, 2003, 09:01:39  
Job time : 54.5887 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:11:38 ; Search time 38.1589 Seconds

(without alignments)  
4403.399 Million cell updates/sec

Title: SEQ4-579-593-14

Perfect score: 1 MELALCRWGLLALLRPGA.....TFKGTPTAENPEYIGLDVPIV 1261

Sequence:

Scoring table:

Gapop 10.0, Gapext 0.5

Searched:

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :	Score	Match	Length	DB	ID	Description
1: /Geneseq_101002.*	6627	96.9	1255	21	AA92620	Human heregulin 2
2: /SID2/gcgdata/geneseq/geneq-emb1/AA1980.DAT.*	6627	96.9	1255	22	AAE12130	Human tyrosine kin
3: /SID2/gcgdata/geneseq/geneq-emb1/AA1981.DAT.*	6627	96.9	1255	22	AAE60167	HER2 transgene pla
4: /SID2/gcgdata/geneseq/geneq-emb1/AA1982.DAT.*	6627	96.9	1255	23	AAU74545	Human HER2 (erbB2)
5: /SID2/gcgdata/geneseq/geneq-emb1/AA1983.DAT.*	6621	96.8	1255	17	AAU01111	HER-2/neu protein.
6: /SID2/gcgdata/geneseq/geneq-emb1/AA1984.DAT.*	6621	96.8	1255	20	AA92406	Human HER-2/neu on
7: /SID2/gcgdata/geneseq/geneq-emb1/AA1985.DAT.*	6621	96.8	1255	21	AAE21198	Human HER-2/neu pr
8: /SID2/gcgdata/geneseq/geneq-emb1/AA1986.DAT.*	6621	96.8	1255	21	AAE84780	Amino acid sequenc
9: /SID2/gcgdata/geneseq/geneq-emb1/AA1987.DAT.*	6621	96.8	1255	22	AAE85458	Human HER-2/neu pr
10: /SID2/gcgdata/geneseq/geneq-emb1/AA1988.DAT.*	6621	96.8	1255	22	AAE88267	HER2/neu amino aci

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

11	6621	96.8	1255	23	AAE24067	Human Her-2 protei
12	6621	96.8	1255	23	AAE20479	Human Her-2/neu pr
13	6621	96.8	1255	23	AAE51143	Human Her-2/neu on
14	6621	96.8	1255	23	AAU77114	Human Her-2/neu po
15	6621	96.8	1255	14	AAE39568	Sequence of c-erbB
16	6457	94.4	1223	23	AAU98923	Human breast cance
17	6304	92.2	1200	21	AAE21200	Human HER-2/neu pr
18	5837.5	85.3	1256	21	AAE21199	Rat HER-2/neu prot
19	5837.5	85.3	1256	23	AAE51144	Mouse Her-2/neu onco
20	5810.5	84.9	1256	22	AAE21206	Mouse Her-2/neu pr
21	5810.5	84.9	1256	22	AAE62860	Amino acid sequenc
22	5810.5	84.9	1256	22	AAE51151	Mouse Her-2/neu on
23	4707	68.8	919	21	AAE21203	Human HER-2/neu fu
24	4707	68.8	919	23	AAE51148	Her-2/neu extracel
25	3660.5	57.9	920	23	AAE51152	Mouse Her-2/neu ex
26	3660.5	57.9	926	23	AAE51153	Mouse Her-2/neu ex
27	3591	52.5	712	21	AAE21204	Human HER-2/neu fu
28	3591	52.5	712	23	AAE51149	Her-2/neu extracel
29	3445	50.4	782	18	AAE19764	Her-2-GN-CGF immuno
30	3443	50.3	653	21	AAE21200	Extracellular HER-
31	3443	50.3	653	23	AAE51145	Human ErbB2 oncopr
32	3405	49.8	645	22	AAE60408	Human ErbB2 extrac
33	3405	49.8	645	22	AAE61593	DC9acFV-erbB2EC fu
34	3340	48.8	951	21	AAE44993	Extracellular port
35	3237	47.3	624	11	AAE08222	Amino acid sequenc
36	3110	45.5	1210	21	AAE19255	Human EGF receptor
37	3110	45.5	1210	21	AAE50616	Human EGF receptor
38	3110	45.5	1210	23	AAE23019	Human epidermal gr
39	3110	45.5	1210	23	AAE50768	Human epidermal gr
40	3108	45.4	1210	22	AAE68420	Amino acid sequenc
41	3084	45.1	583	23	AAE20483	Human protein for
42	3084	45.1	587	23	AAE20481	Human protein for
43	3083	45.1	589	23	AAE20482	Human protein for
44	3083	45.1	600	23	AAE20484	Human protein for
45	3069	44.9	1210	23	ABP51768	Human epidermal gr

#### ALIGNMENTS

RESULT 1	AA92620	standard; Protein; 1255 AA.
ID	AA92620	standard; Protein; 1255 AA.
XX	AA92620;	
AC	AA92620;	
XX	10-AUG-2000	(first entry)
XX	Human heregulin 2 (Her2).	
DE	Human heregulin 2 (Her2).	
XX	Her2; vaccination; cytotoxic T-lymphocyte immunity;	
KW	self-protein; cancer; breast cancer; prostate cancer;	
KW	cell-associated peptide antigen; foreign epitope.	
XX	Homo sapiens.	
OS	Homo sapiens.	
XX	Key	Location/Qualifiers
XX	Key	1..173
FT	Domain	/label= N-terminal
FT	Domain	/note= "mature polypeptide"
FT	Region	5..25
FT	Region	/label= insertion region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	59..73
FT	Region	/label= insertion region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	103..117
FT	Region	/label= insertion region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	149..163
FT	Region	/label= insertion region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Domain	174..323

FT	Region	/label= Cysteine_rich_domain
FT	Region	210..224
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	250..264
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	324..483
FT	Region	/label= ligand_binding_domain
FT	Region	325..339
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	369..383
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	465..479
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	484..623
FT	Region	/label= Cysteine_rich_domain
FT	Region	579..593
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	624..654
FT	Region	/label= Transmembrane_domain
FT	Region	632..652
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	653..667
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	655..1010
FT	Region	/label= Tyrosine_kinase_domain
FT	Region	661..675
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	695..709
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	710..730
FT	Region	/label= insertion_region
FT	Region	/note= "suitable for foreign epitope insertion"
FT	Region	1011..1235
FT	Region	/label= C-terminal_domain
FT	Region	
XX	MO20020027-A2.	
PN	13-APR-2000.	
XX	05-OCT-1999;	99MO-DK00525.
XX	05-OCT-1998;	98DK-0001261.
XX	20-OCT-1998;	98US-0105011.
XX	(MEBI-) M & E BIOTECH AS.	
XX	Steenaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I;	
XX	Gautam A, Birk P, Karlsson G;	
XX	WPI; 2000-349917/30.	
XX	N-PSDB; AAA09455.	
XX	Inducing immune responses to weakly immunogenic, tumor associated	
XX	peptide antigens for the treatment of breast and prostate cancer	
XX	Claim 62; Page 193-198; 220p; English.	
XX	This is the human heregulin 2 (Her2) sequence. Immunogenic analogues	
XX	Her2 can be used in the claimed method as an autovaccine to induce a	
XX	response. Subdominant CTL epitopes, antibody binding regions and	
XX	cysteine residues involved in disulfide bonds are preserved in the	
XX	immunogenized forms. Regions suitable for the insertion of foreign T	
XX	helper epitopes were identified (see features table). The method	

QY	777	AYWAGVSPYVSRLLGICLTSTVOLNOLMPYCGCLDHVENRRLSODLLMCMQIA	836
Db	771	AYWAGVSPYVSRLLGICLTSTVOLNOLMPYCGCLDHVENRRLSODLLMCMQIA	830
QY	837	KGMSYLEDBVLVHRDLAARNVLVXSPNHVKITTFGLARLLDIDETEHADGAKYPIKMA	896
Db	831	KGMSYLEDBVLVHRDLAARNVLVXSPNHVKITTFGLARLLDIDETEHADGAKYPIKMA	890
QY	897	LESLIRRFTHQSDVWSYGYVTWMLMTGAKPYDGIIPARELPDLLEKXERLPORPCTID	956
Db	891	LESLIRRFTHQSDVWSYGYVTWMLMTGAKPYDGIIPARELPDLLEKXERLPORPCTID	950
QY	957	VYMIWVKCMWIDSECRPFRELVSEFSMAADPOCFVVIQNEHDIGPASPILDSTFYRSLLE	1018
Db	951	VYMIWVKCMWIDSECRPFRELVSEFSMAADPOCFVVIQNEHDIGPASPILDSTFYRSLLE	1010
QY	1017	DDDMGDVLDAEEYLVPOQGFPCPDPAFAGAMVHRRHSSSTRSGGGLTLGLEPSEBEA	1076
Db	1011	DDDMGDVLDAEEYLVPOQGFPCPDPAFAGAMVHRRHSSSTRSGGGLTLGLEPSEBEA	1070
QY	1077	PRSLPAPSEAGSGVFDGDLGMAKAGLQSLPTHDPSFLQYSEDPVLPBSETDGYAP	1136
Db	1071	PRSLPAPSEAGSGVFDGDLGMAKAGLQSLPTHDPSFLQYSEDPVLPBSETDGYAP	1130
QY	1137	LTCSPQPEYVQOPDVRPQSPSPRGSLPAAAPAGATLTERATLSPGXGKGVKDVAFEGGA	1196
Db	1131	LTCSPQPEYVQOPDVRPQSPSPRGSLPAAAPAGATLTERATLSPGXGKGVKDVAFEGGA	1190
QY	1197	VENEEYLTPOGGAAPQPHPPAFSPADNLVYWDQDPBERGAPSTFKGTPTAENPELJG	1256
Db	1191	VENEEYLTPOGGAAPQPHPPAFSPADNLVYWDQDPBERGAPSTFKGTPTAENPELJG	1250
QY	1257	LDVPPV 1261	
Db	1251	LDVPPV 1255	

XX	RESULT 2
XX	AAE12130
XX	ID AAE12130 standard; Protein; 1255 AA.
XX	AC AAE12130;
XX	DT 18-DEC-2001 (first entry)
XX	DE Human tyrosine kinase-type receptor, HER-2.
XX	Therapeutic compound; major histocompatibility complex; vaccine;
XX	KM antigenic peptide; MHC; immunoregulatory; immune response; HER-2
XX	KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
XX	KM antigen presenting cell; human; tyrosine kinase-type receptor.
XX	OS Homo sapiens.
XX	Key
XX	FH Location/Qualifiers
XX	FT Region
XX	FT 774..782
XX	/note="Antigenic epitope"
XX	PN WO200168677-A2.
XX	PD 20-SEP-2001.
XX	PF 16-MAR-2001; 2001WO-US40328.
XX	PR 16-MAR-2000; 2000US-0527487.
XX	PA (GEN2 ) GENZYME CORP.
XX	PI Nicolette CA;
XX	WPI, 2001-616284/71.
XX	DR N-PSDB; AAD19731.
XX	DR

PT Novel synthetic therapeutic compound for inducing immune response and  
PT for use in adoptive immunotherapy, has enhanced binding to major  
PT histocompatibility molecules and enhanced immunoregulatory properties  
XX -  
PT  
PT  
XX  
XX  
PT Claim 4; Page 63-67; 69pp; English.

The invention relates to synthetic therapeutic compounds (antigenic peptides) with enhanced binding to major histocompatibility complex (MHC) molecules and enhanced immunoregulatory properties relative to their natural counterparts. Compounds of the invention are useful for inducing an immune response in a subject and for use in adoptive immunotherapy. They are useful as components of anti-cancer vaccines and to expand immune effector cells that are specific for cancers characterized by expression of the breast cancer antigen, HER-2. Polynucleotides that encode peptides of the invention are useful as hybridisation probes and as primers for the detection of genes of gene transcripts that are expressed in antigen presenting cells (APCs), to confirm transduction of polynucleotides into host cells. The present sequence is human tyrosine kinase-type receptor, HER-2. Compounds of the invention are designed based on the HER-2 antigenic peptide (774-782).

SQ	Sequence	1255	AA;
----	----------	------	-----

Query Match	96.9%	Score 6627	DB 22	Length 1255
Best Local Similarity	97.0%	Pred. No. 0		
Matches 1227	Conservative	6	Mismatches 18	Indels 14
				Gaps 2

Qy	MEJALACRMGLLALLPFGAASSTQVCTSDMKMLRPAESPETHMLMHLVQSGCVVGNL	60
Dd	1 MELALCRMGILLALLPFGAASSTQVCTSDMKMLRPAESPETHMLMHLVQSGCVVGNL	60
Qy	ELTYLPLETNASLSLTDIOEVOGVYLIANNOVROVPLORLRIVRGTQVLFEDNYALAVDNG	120
Dd	61 ELTYLPLETNASLSLTDIOEVOGVYLIANNOVROVPLORLRIVRGTQVLFEDNYALAVDNG	120
Qy	DLPLNNTPTVYGASPGGLRELOJRSJTELLKSGVLIQORNPOLCYODTILMKDIFFHKNNQLA	180
Dd	121 DLPLNNTPTVYGASPGGLRELOJRSJTELLKSGVLIQORNPOLCYODTILMKDIFFHKNNQLA	180
Qy	LTLIDPNRSRACHPGSPMKCKSRGCGESSEBDCOSITRTVCAGGACARCKGRLPFDCCHEOC	240
Dd	181 LTLIDPNRSRACHPGSPMKCKSRGCGESSEBDCOSITRTVCAGGACARCKGRLPFDCCHEOC	240
Qy	AAAGCTPRKSDCIACIHFHSIGICELHCPALVTYNTDTFESMPREGRYTFGASCYTACP	300
Dd	241 AAAGCTPRKSDCIACIHFHSIGICELHCPALVTYNTDTFESMPREGRYTFGASCYTACP	300
Qy	YNTLSDVGSCTLVCPRLNQEVTADJDQORCEKSKRCARCYGLCMQYIKANSKFLGIT	360
Dd	301 YNTLSDVGSCTLVCPRLNQEVTADJDQORCEKSKRCARCYGLCMQYIKANSKFLGIT	360
Qy	ELTFPACCKKI FGSILAFLESPFEDGPASNTAFLQEOQLQVEETLEITNGYLIISAMPDSL	420
Dd	361 ELTFPACCKKI FGSILAFLESPFEDGPASNTAFLQEOQLQVEETLEITNGYLIISAMPDSL	420
Qy	DLSPFONLOVINGRILNHGAYSILTQIGIGISWLBLSRLRELGSLALIHNTHLCEVNHV	480
Dd	421 DLSPFONLOVINGRILNHGAYSILTQIGIGISWLBLSRLRELGSLALIHNTHLCEVNHV	480
Qy	PMDOLFERNHQAALLHTANRPEDECYGBEGIACHQICAGHCGMGPGPTQCVVCSQFLRGQEC	540
Dd	481 PMDOLFERNHQAALLHTANRPEDECYGBEGIACHQICAGHCGMGPGPTQCVVCSQFLRGQEC	540
Qy	VEBCRLTQJLPREYVYNNARHCLPCHPEOCOPMGASTGCGPRADQVCANHKDPRFVAC	600
Dd	541 VEBCRLTQJLPREYVYNNARHCLPCHPEOCOPMGASTGCGPRADQVCANHKDPRFVAC	600
Qy	PFNNFTVSEFWLVRPVASAHLE----PDEBGAQCPICINCTHSCVDLJDCKCPAEORASP	656
Dd	601 PFNNFTVSEFWLVRPVASAHLE----PDEBGAQCPICINCTHSCVDLJDCKCPAEORASP	656
Qy	YVFDLSYMPRIKWFPEDEBACQPCPCINCTHSCVDLJDCKCPAEORASP	650
Dd	601 YVFDLSYMPRIKWFPEDEBACQPCPCINCTHSCVDLJDCKCPAEORASP	650

QY 657 LTSIVSAVGLLVVGVFGILIKRQOKIRKRYTMRLLQETELVEPLTPSGAMPNQA 716  
 DB 651 LTSIVSAVGLLVVGVFGILIKRQOKIRKRYTMRLLQETELVEPLTPSGAMPNQA 710  
 QY 717 QMRILKETELRKVKVLSGSAFGTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDE 776  
 DB 711 QMRILKETELRKVKVLSGSAFGTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDE 770  
 QY 777 AYVWAGVSPVSSLLGICLTSTVQLTQMLPFGCLLDHRENRGRGSGDLMNMCQIA 836  
 DB 771 AYVWAGVSPVSSLLGICLTSTVQLTQMLPFGCLLDHRENRGRGSGDLMNMCQIA 830  
 QY 837 KQMSYLEDVRLVHRDLAARVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMA 896  
 DB 831 KQMSYLEDVRLVHRDLAARVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMA 890  
 QY 897 LESTLRRRPTHOSVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITD 956  
 DB 891 LESTLRRRPTHOSVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITD 950  
 QY 957 VYIMVWKWMIIDSECRPFRELVESEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1016  
 DB 951 VYIMVWKWMIIDSECRPFRELVESEFSRMARDPQRFVVIQNEDLGPASPLDSTFYRSLLE 1010  
 QY 1017 DDDMGDLVDAEYVLPQOGFPCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEBEA 1076  
 DB 1011 DDDMGDLVDAEYVLPQOGFPCDPAPAGAGMHHRRSSSTRSGGDLTLGLEPSEBEA 1070  
 QY 1077 PRSPPLASSEGAGSDVDFGDDLGMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGVAP 1136  
 DB 1071 PRSPPLASSEGAGSDVDFGDDLGMAKGLQSLPTHDPSPLQRYSEDPVPLPSETDGVAP 1130  
 QY 1137 LTCSPQPEYVNPQDVRPQPSPREGRLPAARPAQATLERAKTSLPGKNGVYKDVAFGGA 1196  
 DB 1131 LTCSPQPEYVNPQDVRPQPSPREGRLPAARPAQATLERAKTSLPGKNGVYKDVAFGGA 1190  
 QY 1197 VENPEYVLTPOGGAPOHPHPAPAFDNLVYMQDPPERGAPOSTKGTPTAENPEYLG 1256  
 DB 1191 VENPEYVLTPOGGAPOHPHPAPAFDNLVYMQDPPERGAPOSTKGTPTAENPEYLG 1250  
 QY 1257 LDVEV 1261  
 DB 1251 LDVEV 1255

RESULT 3  
 ID AAB60167 standard; Protein; 1255 AA.  
 AC AAB60167;  
 DT 03-APR-2001 (first entry)  
 DE HER2 transgene plasmid construct encoded protein.  
 KM Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;  
 XX antibody.  
 OS Homo sapiens.  
 OS Synthetic.  
 XX WO200100244-A2.  
 XX 04-JAN-2001.  
 XX 23-JUN-2000; 2000WO-US-7229.  
 XX 25-JUN-1999; 99US-0141316.  
 XX 16-MAR-2000; 2000US-0189844.  
 XX (GETH ) GENENTECH INC.  
 XX Erickson S, Schwall R.

XX MPI: 2001-061962/07.  
 DR N-PSDB; AAF24297.  
 XX Treating tumors, particularly breast cancers, which overexpress an ErbB  
 PT receptor and does not respond to an anti-ErbB antibody, comprises  
 PT conjugating the antibody to a maytansinoid -  
 XX  
 PS Example 3; Fig 4; 92p; English.  
 XX  
 CC The present invention provides a method of treating cancer by  
 CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In  
 CC particular, the antibody is directed against ErbB2 (also known as HER2  
 CC and p185neu). The method is particularly useful in the treatment of  
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,  
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.

SQ Sequence 1255 AA;

Query Match 96.9%; Score 6627; DB 22; Length 1255;  
 Best Local Similarity 97.0%; Pred. No. 0;  
 Matches 1227; Conservative 6; Mismatches 18; Indels 14; Gaps 2;

QY 1 METALCRWGLLALPPGAASVYCTGTMTKLRLPASBETHLMDLRHLYGCGVYQGNL 60  
 DB 1 METALCRWGLLALPPGAASVYCTGTMTKLRLPASBETHLMDLRHLYGCGVYQGNL 60  
 QY 61 ELTYLPTNASLFLDIOEVQGYVLIANNOVROVLOLRIVRGQLFEDNYALAVLNG 120  
 DB 61 ELTYLPTNASLFLDIOEVQGYVLIANNOVROVLOLRIVRGQLFEDNYALAVLNG 120  
 QY 121 DEANNTPTVGSAGRELEQLRSLEILKGVLIORNPOLCYODTIMKDIFFKNNQLA 180  
 DB 121 DEANNTPTVGSAGRELEQLRSLEILKGVLIORNPOLCYODTIMKDIFFKNNQLA 180  
 QY 181 LTLIDTNSRACHPCSPKCKSGRCSSESDCOQLTRVYACGCGARCGPLPTDCHEQC 240  
 DB 181 LTLIDTNSRACHPCSPKCKSGRCSSESDCOQLTRVYACGCGARCGPLPTDCHEQC 240  
 QY 241 AAGCTGPRKSDCLALHFNFHSGICELHCPALVYVTFEESMPNREGYTFGASCVTACP 300  
 DB 241 AAGCTGPRKSDCLALHFNFHSGICELHCPALVYVTFEESMPNREGYTFGASCVTACP 300  
 QY 301 YNYLSTDVGSCTLVCPLEHNOEYTAEDGTQRCCKSKPCARVCYGLGMEHLEBRAVTSAN 360  
 DB 301 YNYLSTDVGSCTLVCPLEHNOEYTAEDGTQRCCKSKPCARVCYGLGMEHLEBRAVTSAN 360  
 QY 361 ELEFAGCKKIRGSLAFLESPFDGPASTAPLQBPQLQVETLEETITGVLYISAMPDSL 420  
 DB 361 IOEFAGCKKIRGSLAFLESPFDGPASTAPLQBPQLQVETLEETITGVLYISAMPDSL 420  
 QY 421 DLSEFQNTQVIRGRILHNGAVSLTLQGLIGISWLGRLSRLRELSGIALIHNTHLCFVATV 480  
 DB 421 DLSEFQNTQVIRGRILHNGAVSLTLQGLIGISWLGRLSRLRELSGIALIHNTHLCFVATV 480  
 QY 481 PWDQLFKNPHQALLHTANRPDECEYGBELACHOLCARHCGPFGTQCVNSQFLRGEC 540  
 DB 481 PWDQLFKNPHQALLHTANRPDECEYGBELACHOLCARHCGPFGTQCVNSQFLRGEC 540  
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECOPOKGSVTCFGEADOCVACHYADPPCVAR 600  
 DB 541 VEECRVLOGLPREYVNAHCLPCHPECOPOKGSVTCFGEADOCVACHYADPPCVAR 600  
 QY 601 PENNFTVFWLRVPRVSAHLE---PDEGACQPCPINCSTHSCVDLDDKQCPAEPORAP 656  
 DB 601 PGG-----VPRDLSYMPIMKFPDEGACQPCPINCSTHSCVDLDDKQCPAEPORAP 650  
 QY 657 LTSIVSAVGLLVVGVFGILIKRQOKIRKRYTMRLLQETELVEPLTPSGAMPNQA 716  
 DB 651 LTSIVSAVGLLVVGVFGILIKRQOKIRKRYTMRLLQETELVEPLTPSGAMPNQA 710  
 QY 717 QMRILKETELRKVKVLSGSAFGTYKGIWIPDGENVKIPVAIKVLRNTSPKANKELDE 776

Db 711 QMRILKETEIRKVKVLSGAFGVYKGIPIPDGENVKIPVAIVLENTSPKANKELDE 770  
 Qy 777 AYVAGVSPYVSRLLIGICTSTVOLVQLMPYGCLLDHYENRGRIGSODLLNMCQIA 836  
 Db 771 AYVAGVSPYVSRLLIGICTSTVOLVQLMPYGCLLDHYENRGRIGSODLLNMCQIA 830  
 Qy 837 KGMSTYEDVLYHRDLAARVLYKSPNHVKITDPGLARLDIDETEHADGKVPIKMA 896  
 Db 831 KGMSTYEDVLYHRDLAARVLYKSPNHVKITDPGLARLDIDETEHADGKVPIKMA 890  
 Qy 897 LESLRERFTHOSDWSYGTWELMTFGAKPYDGIIPAREIPDLLEKGBELPQPICTID 956  
 Db 891 LESLRERFTHOSDWSYGTWELMTFGAKPYDGIIPAREIPDLLEKGBELPQPICTID 950  
 Qy 957 VYIMVYKCMWIDSECRPFRELVESESRMARDQRFVYIIONEDLGAPSLDSTFYRSLLE 1016  
 Db 951 VYIMVYKCMWIDSECRPFRELVESESRMARDQRFVYIIONEDLGAPSLDSTFYRSLLE 1010  
 Qy 1017 DDDMGDLVDAEELVPPQGFPCDPAPAGAGMHHRRSSSTSGGDLTLIGLEPSEEA 1076  
 Db 1011 DDDMGDLVDAEELVPPQGFPCDPAPAGAGMHHRRSSSTSGGDLTLIGLEPSEEA 1070  
 Qy 1077 PRSPLAPSEGAGSDVFDGDLGMAAGLQSLPTHDPSPLORYSEDETVLPSETDGYAP 1136  
 Db 1071 PRSPLAPSEGAGSDVFDGDLGMAAGLQSLPTHDPSPLORYSEDETVLPSETDGYAP 1130  
 Qy 1137 LTGSPQPEYVNOVDYRPPSPREGLPAPAPAGATLERAKTISPQNGVYKDVAFAGGA 1196  
 Db 1131 LTGSPQPEYVNOVDYRPPSPREGLPAPAPAGATLERAKTISPQNGVYKDVAFAGGA 1190  
 Qy 1197 VENPEYLTPOGGAAPRPPAPSPAFDNLXYWDOPPERGAPSPSTFKGPTAENPEYLG 1256  
 Db 1191 VENPEYLTPOGGAAPRPPAPSPAFDNLXYWDOPPERGAPSPSTFKGPTAENPEYLG 1250  
 Qy 1257 LDVPV 1261  
 Db 1251 LDVPV 1255

RESULT 4  
 AAU74545  
 ID AAU74545 standard, Protein, 1255 AA.  
 AC AAU74545;  
 DT 23-APR-2002 (first entry)  
 DE Human HER2 (ErbB2) polypeptide.  
 KW Human; HER2; ErbB; epidermal growth factor receptor; receptor;  
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;  
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;  
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;  
 KW glial disorder; astrocytal disorder; hypothalamic disorder;  
 KW glandular disorder; macrophagal disorder; epithelial disorder;  
 KW stromal disorder; blastococelic disorder; inflammatory disorder;  
 KW angiogenic disorder; immunological disorder.  
 OS Homo sapiens.  
 PN US2002001587-A1.  
 PD 03-JAN-2002.  
 PF 16-MAR-2001; 2001US-0811123.  
 PR 16-MAR-2000; 2000US-189844P.  
 PR 05-OCT-2000; 2000US-238327P.  
 PA (ERIC/) ERICKSON S.  
 PA (SCHW/) SCHWALL R.  
 PA (SLIM/) SLIKOWSKI M.  
 XX

PI Erickson S, Schwall R, Slikowski M;  
 XX WPI; 2002-163686/21.  
 DR N-PSDB; ABK14058.  
 XX  
 PT Treating tumour characterised by overexpression of epidermal growth  
 PT factor receptor, ErbB or cancer in mammal, comprises administering  
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal  
 PS  
 XX Example 3; Fig 7; 93pp; English.  
 CC The invention relates to treating a tumour in a mammal, where the tumour  
 CC is characterised by the overexpression of an epidermal growth factor  
 CC receptor (ErbB) and does not respond or responds poorly, to treatment  
 CC with an anti-ErbB antibody, comprising administering to the mammal an  
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for  
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,  
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,  
 CC prostate and bladder, preferably breast cancer. The breast cancer is a  
 CC metastatic breast cancer or an aggressive form of metastatic breast  
 CC cancer which overexpresses ErbB2. The method is also useful for treating  
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,  
 CC epithelial, stromal, blastococelic, inflammatory, angiogenic and  
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)  
 CC polypeptide of the invention.  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 96.9%; Score 6627; DB 23; Length 1255;  
 Best Local Similarity 97.0%; Pred. No. 0;  
 Matches 1227; Conservative 6; Mismatches 18; Indels 14; Gaps 2;  
 Qy 1 MELAALCRWGLLALLPPGAASVQCTGTDKRLPASPETHLMDRLYQCGQVQGNL 60  
 Db 1 MELAALCRWGLLALLPPGAASVQCTGTDKRLPASPETHLMDRLYQCGQVQGNL 60  
 Qy 61 ELTYLPTNASLSPIDIOEVOGYVLIANOVROYPLQRLIVRGTLQFEDNYALAVLNG 120  
 Db 61 ELTYLPTNASLSPIDIOEVOGYVLIANOVROYPLQRLIVRGTLQFEDNYALAVLNG 120  
 Qy 121 DPLNNTTPVTGASPGGLRELQRLSTELKSGVLIQRNPQLCYODTILMKDIFHNKOLA 180  
 Db 121 DPLNNTTPVTGASPGGLRELQRLSTELKSGVLIQRNPQLCYODTILMKDIFHNKOLA 180  
 Qy 181 LTLIDTNSRACHPSPCKSKRCWGBSSSDCCSLTRTVACAGCARCKGPIPTDCHEQC 240  
 Db 181 LTLIDTNSRACHPSPCKSKRCWGBSSSDCCSLTRTVACAGCARCKGPIPTDCHEQC 240  
 Qy 241 AAGCTGPRGSDCLCLHFNHSGICEHLHPALVTYNTDFEEMPEBEGYTGACACVYACP 300  
 Db 241 AAGCTGPRGSDCLCLHFNHSGICEHLHPALVTYNTDFEEMPEBEGYTGACACVYACP 300  
 Qy 301 YNYLSTDVGSCTVLCPLHNOEVTAEADGTORCEKSKPCARVYCYGLQMOYIKANSKFIGIT 360  
 Db 301 YNYLSTDVGSCTVLCPLHNOEVTAEADGTORCEKSKPCARVYCYGLQMOYIKANSKFIGIT 360  
 Qy 361 ELEPAGCKKIGSLAFLPESFDGDPASNTAPLOEQLQVFTLEITGYLYISAMPDPLP 420  
 Db 361 IGFAPGCKKIGSLAFLPESFDGDPASNTAPLOEQLQVFTLEITGYLYISAMPDPLP 420  
 Qy 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLRLRELGSGLALIHNNTHLCFHTV 480  
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLIGISWLGRLRLRELGSGLALIHNNTHLCFHTV 480  
 Qy 481 PMDQLFRNPQALHTANRPEDECVGSLACHQLCARGCKGPGPTCCVNCSCPLRQEC 540  
 Db 481 PMDQLFRNPQALHTANRPEDECVGSLACHQLCARGCKGPGPTCCVNCSCPLRQEC 540  
 Qy 541 VEECRVYQGLPREYVNAARCLPCHPECOPONGSVTCGPEADOCVCAAHYKDPFCVARC 600  
 Db 541 VEECRVYQGLPREYVNAARCLPCHPECOPONGSVTCGPEADOCVCAAHYKDPFCVARC 600  
 Qy 601 PFNNFTVSWLVRPKVASHSLE---DDEGACQPCPINCSTHSCVDLDDKGCAPABRASP 656



```

Db 601 PSQ-----VKEDLSYMPIMKFPDEEGACQPCPCINTHSCVLDLKGCAEGRASP 650
Qy 657 LTIIVAVGVGILLVVLGVFGILIKRQOKIRKYMRLLOETLVEPLTPSGAMPNOA 716
Db 651 LTIIVAVGVGILLVVLGVFGILIKRQOKIRKYMRLLOETLVEPLTPSGAMPNOA 710
Qy 717 QMEILKETELRKVKVGGSAFGTVYVGIWIPDEENKIPAIIVLEENISPKXKEILDE 776
Db 711 QMEILKETELRKVKVGGSAFGTVYVGIWIPDEENKIPAIIVLEENISPKXKEILDE 770
Qy 777 AYMAGVGSFYSRLLGLCTSTVOLTQMLPYGCLLDHYENRGRGLSGDILNMCQIA 836
Db 771 AYMAGVGSFYSRLLGLCTSTVOLTQMLPYGCLLDHYENRGRGLSGDILNMCQIA 830
Qy 837 KGSYSLEDAVLNHRDLAARVLYKSPNHVKITDFGLARLIDIDETEHADGGKVPKMA 896
Db 831 KGSYSLEDAVLNHRDLAARVLYKSPNHVKITDFGLARLIDIDETEHADGGKVPKMA 890
Qy 897 LESLRRRFTHQSDVMSGYTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTID 956
Db 891 LESLRRRFTHQSDVMSGYTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPICTID 950
Qy 957 VYIMVWKMMIDSECRPFRELVSSESRMARDPQRFVYIQNEDLGSPASPLDSTFYRSLLE 1016
Db 951 VYIMVWKMMIDSECRPFRELVSSESRMARDPQRFVYIQNEDLGSPASPLDSTFYRSLLE 1010
Qy 1017 DDMGDLVDAEYLVPOQGFPCPDPAAGAGMHHRRSSSTSGGGDLTLIGPSEEEA 1076
Db 1011 DDMGDLVDAEYLVPOQGFPCPDPAAGAGMHHRRSSSTSGGGDLTLIGPSEEEA 1070
Qy 1077 PRSLPASEGAGSDVDFDGLGMAKAGLQSLPTHDSPLQRYSEDPVPLPSETDGYVAP 1136
Db 1071 PRSLPASEGAGSDVDFDGLGMAKAGLQSLPTHDSPLQRYSEDPVPLPSETDGYVAP 1130
Qy 1137 LTGSPQPEYVNOQDVVRQPPSPRRGPIPARPAGATLEAKTISPGKNGVYKVFAPGGA 1196
Db 1131 LTGSPQPEYVNOQDVVRQPPSPRRGPIPARPAGATLEAKTISPGKNGVYKVFAPGGA 1190
Qy 1197 VENPEYLTPOGGAAPQPPPPAPSPAFNDLYWDDPPEPAGAPSTFKGPTAENEYILG 1256
Db 1191 VENPEYLTPOGGAAPQPPPPAPSPAFNDLYWDDPPEPAGAPSTFKGPTAENEYILG 1250
Qy 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 5
AAW01111
ID AAW01111 standard; Protein; 1255 AA.
XX
AC AAW01111;
XX
DT 01-JAN-1997 (first entry)
XX
DE HER-2/neu protein.
XX
KW HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
KW breast cancer; ovary cancer; colon cancer; lung cancer;
KW prostate cancer; immunisation; tumour; vaccine; vector.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FH 676..1255
FT Domain /label= Intracellular domain
FT /note="Claimed domain, useful for immunisation"
XX
PN M0630514-A1.
XX
PD 03-OCT-1996.

```

```

PF 28-MAR-1996; 96MO-US01689.
XX
PR 31-MAR-1995; 95US-0414417.
XX
XX (UNIM) UNIV WASHINGTON.
XX
PI Cheever MA, Disis ML;
XX
DR WPI. 1996-455361/45.
DR N-PSDB; AAT40739.
XX
PT DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
PT treatment of malignancies with which the HER-2/neu oncogene is
PT associated
XX
PS Claim 2; Page 56-61; 71pp; English.
XX
XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;
XX
Query Match 96.8%; Score 6621; DB 17; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;
Qy 1 METALCWGILLALPPGASTQCTGDMKRLPASPETHLMLRLLYGGCCVYQCNL 60
Db 1 METALCWGILLALPPGASTQCTGDMKRLPASPETHLMLRLLYGGCCVYQCNL 60
Qy 61 ETTYLPTNASLFLDIOEVQGYVLIANOVQVPLQRLIVRGTLFEDNYALAVLDNG 120
Db 61 ETTYLPTNASLFLDIOEVQGYVLIANOVQVPLQRLIVRGTLFEDNYALAVLDNG 120
Qy 121 PLNNTTPTVGTASPGGLRELQRLSLTELKGVVLIQRPOLCYDTILMKDIFHNQOLA 180
Db 121 PLNNTTPTVGTASPGGLRELQRLSLTELKGVVLIQRPOLCYDTILMKDIFHNQOLA 180
Qy 181 LTLIDNRSRACHPSPCKSGRCGSESEDCOSLTRVCAAGGRCRGPPTDCHEQC 240
Db 181 LTLIDNRSRACHPSPCKSGRCGSESEDCOSLTRVCAAGGRCRGPPTDCHEQC 240
Qy 241 AAGCTGPKHSDCLACIENHSGICELCPALVTYNTDIFESMPNDEGRYTGACSVTACP 300
Db 241 AAGCTGPKHSDCLACIENHSGICELCPALVTYNTDIFESMPNDEGRYTGACSVTACP 300
Qy 301 YNYLSTDVSGCTLVPLNNOEVTAEADGTQRCBEKSKPCARVYCYGLGMXYTANSKFIGIT 360
Db 301 YNYLSTDVSGCTLVPLNNOEVTAEADGTQRCBEKSKPCARVYCYGLGMXYTANSKFIGIT 360
Qy 361 ELEFAGCKKIFGSLAFLESPFGDPASNTALQPEQLQVFETLEETGYLISAMPDLSLP 420
Db 361 IOEFAGCKKIFGSLAFLESPFGDPASNTALQPEQLQVFETLEETGYLISAMPDLSLP 420
Qy 421 DLSVFQNTQVTRGLIANGAYSLTQGLGISWLGSLRLSGSLALIHNTHLCEVHTV 480
Db 421 DLSVFQNTQVTRGLIANGAYSLTQGLGISWLGSLRLSGSLALIHNTHLCEVHTV 480
Qy 481 PMDOLFRRPHQALLHTARPEDECEGELACHOICARHCGMGPOTOCVNSOFLRGEC 540
Db 481 PMDOLFRRPHQALLHTARPEDECEGELACHOICARHCGMGPOTOCVNSOFLRGEC 540
Qy 541 VEECRVILQGLPREYVNAHCHLPCHPECCQPNQSGVTCFQPEADQCVACAHYDPPFCVARC 600
Db 541 VEECRVILQGLPREYVNAHCHLPCHPECCQPNQSGVTCFQPEADQCVACAHYDPPFCVARC 600

```

QY 601 PFNNFTVSWLRPKYSASHLE----PDEGACQPCPINCTHSCVDLDKGCBAEQAS 656  
 Db 601 PSQ-----VKPDLSTWPIWKFPDEGACQPCPINCTHSCVDLDKGCBAEQAS 650  
 QY 657 LTVSAVVGILLVWVGVVFGILLIKRROCKIRKRYTRRLLOETELVEPIPTSGANPNA 716  
 Db 651 LTVSIAVVGILLVWVGVVFGILLIKRROCKIRKRYTRRLLOETELVEPIPTSGANPNA 710  
 QY 717 QMRILKETELRKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLENTSPKANKELDE 776  
 Db 711 QMRILKETELRKVKVLSGAFGTYYKGIWIPDGENVKIPVAIKVLENTSPKANKELDE 770  
 QY 777 AAYMAGVSPYVRLGICLTSTVQVLTQCLMPYGCILDHRENRGLSGDILNMCQIA 836  
 Db 771 AAYMAGVSPYVRLGICLTSTVQVLTQCLMPYGCILDHRENRGLSGDILNMCQIA 830  
 QY 837 KGMSTYLEDVRLVHRDLAARNVLYKSPHVKITDFGLARLIDETEHADGCKVPIKMA 896  
 Db 831 KGMSTYLEDVRLVHRDLAARNVLYKSPHVKITDFGLARLIDETEHADGCKVPIKMA 890  
 QY 897 LESTLRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITD 956  
 Db 891 LESTLRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITD 950  
 QY 957 VYIMVYKCMWIDSECRPRFRELVSFESRMARDPQRFVITQNEDLGFPASPLDSTFYRSLLE 1016  
 Db 951 VYIMVYKCMWIDSECRPRFRELVSFESRMARDPQRFVITQNEDLGFPASPLDSTFYRSLLE 1010  
 QY 1017 DDMGMDLVDAEYLVPOGFFCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEBEA 1076  
 Db 1011 DDMGMDLVDAEYLVPOGFFCPDPAAGAGVHHRSSSTRSGGDLTLGLEPSEBEA 1070  
 QY 1077 PRSPLAPBSGAGSDVDGDLGMAAGKLOSLPTHDSPLQRYSEDTVPPLPSTDDGVAP 1136  
 Db 1071 PRSPLAPBSGAGSDVDGDLGMAAGKLOSLPTHDSPLQRYSEDTVPPLPSTDDGVAP 1130  
 QY 1137 LTCSPQPEYVNDPVAPOPPSPREGPLPAARPAATLIERAKTILSPKNGVYKDVAFGGA 1196  
 Db 1131 LTCSPQPEYVNDPVAPOPPSPREGPLPAARPAATLIERAKTILSPKNGVYKDVAFGGA 1190  
 QY 1197 VENPEYLTLPQGAAPQHPAPSPAFDNLVYWDQDPPRGAAPSTFKGPTAENPEYLG 1256  
 Db 1191 VENPEYLTLPQGAAPQHPAPSPAFDNLVYWDQDPPRGAAPSTFKGPTAENPEYLG 1250  
 QY 1257 LDVPV 1261  
 Db 1251 LDVPV 1255

RESULT 6  
 AAM92406 standard; Protein; 1255 AA.  
 XX AAM92406;

AC AAM92406;  
 XX 21-APR-1999 (first entry)  
 DT 21-APR-1999 (first entry)  
 XX Human HER-2/neu oncogene protein.  
 DE Human HER-2/neu oncogene protein.  
 XX HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
 KM malignancy; treatment; tumour.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 XX Key  
 FH Region Location/Qualifiers  
 FT 676..1255  
 FT /note="Region which elicits immune response"  
 XX US5869445-A.  
 XX 09-FEB-1999.  
 PD 01-APR-1996; 96US-0625101.  
 PF

XX 01-APR-1996; 96US-0625101.  
 PR 17-MAR-1993; 93US-0033644.  
 PR 12-AUG-1993; 93US-0106112.  
 PR 31-MAR-1995; 95US-0414417.  
 XX (UNITW) UNITV WASHINGTON.  
 XX Cheever MA, Disis ML;  
 PI MPI: 1999-152835/13.  
 DR N-PSDB; AAX01912.  
 CC Use of HER-2/neu polypeptides - for eliciting an immune response to  
 CC an HER-2/neu associated malignancy, particularly for treating or  
 CC preventing tumours  
 CC Claim 3: Column 31-38; 26pp; English.  
 CC This sequence represents the human HER-2/neu oncogene protein. A fragment  
 CC of this protein is used in a method for eliciting or enhancing an immune  
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and  
 CC B cells to produce an immune response to the HER-2/neu protein. The  
 CC method can be used for immunisation against a malignancy in which the  
 CC HER-2/neu oncogene is associated and in the treatment of an existing  
 CC tumour, or to prevent tumour occurrence or recurrence.  
 CC  
 CC Sequence 1255 AA;  
 CC  
 CC Query Match 96.8%; Score 6621; DB 20; Length 1255;  
 CC Best Local Similarity 96.8%; Pred. No. 0;  
 CC Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;  
 QY 1 MELALCRWGLLALLPPGAASVQCTGTDMLRLPASPTHLDMLRHLYOCCVVGQNL 60  
 Db 1 MELALCRWGLLALLPPGAASVQCTGTDMLRLPASPTHLDMLRHLYOCCVVGQNL 60  
 QY 61 ELTYLPTNASSFLDIOEVQGYVLIANNVRCVPLQRLVYRGTLQFEDNVALAVLDNG 120  
 Db 61 ELTYLPTNASSFLDIOEVQGYVLIANNVRCVPLQRLVYRGTLQFEDNVALAVLDNG 120  
 QY 121 DPLNNTTAVTGASPGELRELQRLSTELTKGAVLIQRNPOLCYOQDTIMKDI FHNQOLA 180  
 Db 121 DPLNNTTAVTGASPGELRELQRLSTELTKGAVLIQRNPOLCYOQDTIMKDI FHNQOLA 180  
 QY 121 DPLNNTTAVTGASPGELRELQRLSTELTKGAVLIQRNPOLCYOQDTIMKDI FHNQOLA 180  
 Db 121 DPLNNTTAVTGASPGELRELQRLSTELTKGAVLIQRNPOLCYOQDTIMKDI FHNQOLA 180  
 QY 181 LTLIDTNRSAHPSPCKGSRWCSESSDQSLRTVCAAGCARCKGRLPTDCHEQC 240  
 Db 181 LTLIDTNRSAHPSPCKGSRWCSESSDQSLRTVCAAGCARCKGRLPTDCHEQC 240  
 QY 181 LTLIDTNRSAHPSPCKGSRWCSESSDQSLRTVCAAGCARCKGRLPTDCHEQC 240  
 Db 181 LTLIDTNRSAHPSPCKGSRWCSESSDQSLRTVCAAGCARCKGRLPTDCHEQC 240  
 QY 241 AAGCTGPRGSDCLAHFNHSGICELHPALVYNTDTFESMPNPEGVYTGASCVTACP 300  
 Db 241 AAGCTGPRGSDCLAHFNHSGICELHPALVYNTDTFESMPNPEGVYTGASCVTACP 300  
 QY 301 YNYLSTDVGSCVLPNCEVTAEDGTQRCCKSPCARCYGLGMQYIRANSKFITG 360  
 Db 301 YNYLSTDVGSCVLPNCEVTAEDGTQRCCKSPCARCYGLGMQYIRANSKFITG 360  
 QY 361 ELFRAGCKKIFGSLAFIPESPDPASNTAPLQEPOLQVFTLEITGYLYISAMPDLP 420  
 Db 361 IOFRAGCKKIFGSLAFIPESPDPASNTAPLQEPOLQVFTLEITGYLYISAMPDLP 420  
 QY 421 DLSVFQNLQVIRGRILNHGAYSLTLOGIGISWLBLSRLRELGSSGLALIHNTTHLCFVHTV 480  
 Db 421 DLSVFQNLQVIRGRILNHGAYSLTLOGIGISWLBLSRLRELGSSGLALIHNTTHLCFVHTV 480  
 QY 481 FWDOLFNNPQALLHTANRPEDCEVGEGLACHQICARHGCGPPTQCVNCSQFLRQEC 540  
 Db 481 FWDOLFNNPQALLHTANRPEDCEVGEGLACHQICARHGCGPPTQCVNCSQFLRQEC 540  
 QY 541 VEEGRVLOGLPREYVNAHCLPCHPECOQONGSVTCGPEADQCVACAANKDPFCVARC 600  
 Db 541 VEEGRVLOGLPREYVNAHCLPCHPECOQONGSVTCGPEADQCVACAANKDPFCVARC 600



```

Db 601 PSQ-----VKPDLSTYPMKPFDEEGACOPCPINCHSCVDLDKGPACQASR 650
QY 657 LSTIVSAVGLILVYVGVFGLIKRQOKIRKYMRLLOETELVEPIBPGAMPNQ 716
Db 651 LSTIVSAVGLILVYVGVFGLIKRQOKIRKYMRLLOETELVEPIBPGAMPNQ 710
QY 717 QMBILKTELKRVKVLGSGAFGVYKGIWIPDENVKIPVAIVLENTSPKANKILDE 776
Db 711 QMBILKTELKRVKVLGSGAFGVYKGIWIPDENVKIPVAIVLENTSPKANKILDE 770
QY 777 AYVWAGVSPYVSRLLGICLTSTVQLTQMLPYGCLLDVHNRGRIGSQDLLNMCQIA 830
Db 771 AYVWAGVSPYVSRLLGICLTSTVQLTQMLPYGCLLDVHNRGRIGSQDLLNMCQIA 830
QY 837 KGMSTLEDAVLVHRLDAARVLYKSPNHVKITDPGLARLDDIDETEHADGKVPKMA 896
Db 831 KGMSTLEDAVLVHRLDAARVLYKSPNHVKITDPGLARLDDIDETEHADGKVPKMA 890
QY 897 LESLRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTID 956
Db 891 LESLRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTID 950
QY 957 VVMIMVCMWIDSECRFRELVSFSPRMARDQRFVYIQNEPLGSPASPLDSTFYSLLE 1016
Db 951 VVMIMVCMWIDSECRFRELVSFSPRMARDQRFVYIQNEPLGSPASPLDSTFYSLLE 1010
QY 1017 DDMGDLVDAEEYLPQGGFFCDDPAFAGAGMHHRRSSSTSGGDLTLGLEPSEEA 1076
Db 1011 DDMGDLVDAEEYLPQGGFFCDDPAFAGAGMHHRRSSSTSGGDLTLGLEPSEEA 1070
QY 1077 PRSPFLABEGAGSDVFPDGLGMAKAGLQSLPTHDPSPFLQRYSEDTPVLPSTSDGYAP 1136
Db 1071 PRSPFLABEGAGSDVFPDGLGMAKAGLQSLPTHDPSPFLQRYSEDTPVLPSTSDGYAP 1130
QY 1137 LTCSPQPEYVNOVDVPRQPPSPREGPLPAAPAGATLTERAKTLSPKNGVYKVPFAGA 1196
Db 1131 LTCSPQPEYVNOVDVPRQPPSPREGPLPAAPAGATLTERAKTLSPKNGVYKVPFAGA 1190
QY 1197 VENPEYLTPOGGAAPQHPAPSPAFEDNLVYWDODPPERKAGPSTFKGTLANDEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPAPSPAFEDNLVYWDODPPERKAGPSTFKGTLANDEYLG 1250
QY 1257 LDVPPV 1261
Db 1251 LDVPPV 1255

```

RESULT 8  
 AAY84780  
 ID AAY84780 standard; Protein; 1255 AA.

AC AAY84780;  
 DT 08-AUG-2000 (first entry)  
 XX  
 DE Amino acid sequence of the SPICE erbB-2 receptor protein.  
 XX  
 KW SPICE erbB-2 receptor protein; cell transformation disorder; cancer;  
 KW tumor cell proliferation; tissue degeneration; arthropathy;  
 KW bone resorption; inflammatory disease; degenerative disorder;  
 KW wound healing.  
 OS Homo sapiens.  
 XX  
 XX WO2000020579-A1.  
 XX  
 XX 13-APR-2000.  
 XX  
 XX 01-OCT-1999; 99WO-CA00912.  
 XX  
 XX 02-OCT-1998; 98US-0165192.  
 XX

```

PA (UYN-) UNIV MCMMASTER.
XX
PI Muller WJ, Siegel PM;
XX WPI, 2000-303768/26.
DR N-PSDB; AAA14812.
XX
PT Nucleic acid encoding an erbB 2 receptor protein designated SPICE
XX erbB-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS Claim 3; Fig 2; 60pp; English.
XX
CC The present sequence represents a SPICE erbB-2 receptor protein. The
CC protein has an in-frame deletion of 16 amino acids, 2 of which are
CC conserved cysteine residues, compared to the unspliced protein. The
CC erbB-2 polynucleotide is used to construct probes for detecting
CC disorders of cell transformation such as cancer. Antibodies to the
CC protein may be used to detect SPICE erbB-2 in a sample. Agents
CC (e.g. antisense oligonucleotides) which inhibit the expression of
CC SPICE erbB-2 are useful for reducing tumor cell proliferation and
CC treating cancer. Substances which stimulate SPICE erbB-2 are useful
CC for treating conditions involving damaged cells including conditions
CC in which degeneration of tissue occurs, such as arthropathy, bone
CC resorption, inflammatory diseases, degenerative disorders of the
CC central nervous system and wound healing.
XX
SQ Sequence 1255 AA;
Query Match 96.8%; Score 6621; DB 21; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;
QY 1 MELALCRWGLLALLPPGAATGYCTGDMKLRLPASPEHLDWLRLHYOGQVYQGNL 60
Db 1 MELALCRWGLLALLPPGAATGYCTGDMKLRLPASPEHLDWLRLHYOGQVYQGNL 60
QY 61 ELTYVPTNASSFLDIOEVQGYVLAHNOVQVPLQSLRIYRGTQLFEDNVALAVLNG 120
Db 61 ELTYVPTNASSFLDIOEVQGYVLAHNOVQVPLQSLRIYRGTQLFEDNVALAVLNG 120
QY 121 PLNNTTPTVGTASPGRLRELQRLSTELIKGVALIQNPOLCYQDTILMKOIFHNQOLA 180
Db 121 PLNNTTPTVGTASPGRLRELQRLSTELIKGVALIQNPOLCYQDTILMKOIFHNQOLA 180
QY 122 PLNNTTPTVGTASPGRLRELQRLSTELIKGVALIQNPOLCYQDTILMKOIFHNQOLA 180
Db 122 PLNNTTPTVGTASPGRLRELQRLSTELIKGVALIQNPOLCYQDTILMKOIFHNQOLA 180
QY 181 LTLIDTNSRACHPCSPCKGSRGWESSDDQSLTRIVCAGGACRCRPLPTDCHQC 240
Db 181 LTLIDTNSRACHPCSPCKGSRGWESSDDQSLTRIVCAGGACRCRPLPTDCHQC 240
QY 241 AAGCTGPRGSDCLAHFNSGICELHCPALVTYNTDTFESMPNEGRYTGASCVTACP 300
Db 241 AAGCTGPRGSDCLAHFNSGICELHCPALVTYNTDTFESMPNEGRYTGASCVTACP 300
QY 301 YNYLSTDVSGCTVCLPCLNOEVTAEADGQRCCKSKPCARVCYGLGMEHLEVRATVSAN 360
Db 301 YNYLSTDVSGCTVCLPCLNOEVTAEADGQRCCKSKPCARVCYGLGMEHLEVRATVSAN 360
QY 361 ELEFAGCKKIFGSLAFPESEFDGPASNTAPLQEBOLQVPELLEITGYLYISAMPDLP 420
Db 361 IOEFAGCKKIFGSLAFPESEFDGPASNTAPLQEBOLQVPELLEITGYLYISAMPDLP 420
QY 421 DLSVFNQNLQVIRGRIHNGAVSLTLOGIGISLWGLRSRLREGSGALIHNTHICFVHTV 480
Db 421 DLSVFNQNLQVIRGRIHNGAVSLTLOGIGISLWGLRSRLREGSGALIHNTHICFVHTV 480
QY 481 PMDOLFRNPHQALLHTANRPEDECVGEELACHQLCARGHCWGPPTOCVNSQPLRGEC 540
Db 481 PMDOLFRNPHQALLHTANRPEDECVGEELACHQLCARGHCWGPPTOCVNSQPLRGEC 540
QY 541 VEECRVULQGLFREYVNAHCLPCHBECOPONGSVTCTGPEADQCVAACHYKDPFVCAR 600
Db 541 VEECRVULQGLFREYVNAHCLPCHBECOPONGSVTCTGPEADQCVAACHYKDPFVCAR 600
QY 601 PFNNFTVSWFLRVKVASHLE----PDSEGAQCPCTINCHSCVDLDKGPACQASR 656

```

```

Db 601 PSG-----VKPLSTMPIMKFPDEGACQPCPINCTHSCVDLDDGCPAEQSPASD
Qy 657 LTSIVSAVGIILVVVLGVFGILIKRROOKIRKTYMRLLQETELVEPLTPSGAMPNQA
Db 651 LTSIISAVGILLVVVLGVFGILIKRROOKIRKTYMRLLQETELVEPLTPSGAMPNQA
Qy 717 QMRILKTELKRYKVLGSSAFGTYVKIWIIPDGENVKIPIAIVLRENTSPKANKELDE
Db 711 QMRILKTELKRYKVLGSSAFGTYVKIWIIPDGENVKIPIAIVLRENTSPKANKELDE
Qy 777 AYVWAGVSEPVSSLLICLTSTVQLTQCLMPYGCCLLDHRENRGRGLSGDILNMCQIA
Db 771 AYVWAGVSEPVSSLLICLTSTVQLTQCLMPYGCCLLDHRENRGRGLSGDILNMCQIA
Qy 837 KGMVLEDEVRLVHRDLAARNVLVKSNNHVKITDFGLARLIDIDETEHADGKVPKMA
Db 831 KGMVLEDEVRLVHRDLAARNVLVKSNNHVKITDFGLARLIDIDETEHADGKVPKMA
Qy 897 LESTLRRRFTHOSQVWSYGVTVWELMTFGAKPYOGIPAREIPDLLEKGERLPQPICTID
Db 891 LESTLRRRFTHOSQVWSYGVTVWELMTFGAKPYOGIPAREIPDLLEKGERLPQPICTID
Qy 957 VYMIWVKCMMIDSECRPFRELVESESRMARDPQRFVYIQNEDLGAPSLDSTFYRSLLE
Db 951 VYMIWVKCMMIDSECRPFRELVESESRMARDPQRFVYIQNEDLGAPSLDSTFYRSLLE
Qy 1017 DDDMGDLVDAEEYLVPOQGFPCPDPAAGAGMHHRRSSSTSGGDLTLGLEPSEEEA
Db 1011 DDDMGDLVDAEEYLVPOQGFPCPDPAAGAGMHHRRSSSTSGGDLTLGLEPSEEEA
Qy 1077 PRSPPLASEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAP
Db 1071 PRSPPLASEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGYVAP
Qy 1137 LTGSPQPEYVNOQPVVRQPPSPREGPLPAAPPAATLTERAKTISPGKNGYVVKVFAFGA
Db 1131 LTGSPQPEYVNOQPVVRQPPSPREGPLPAAPPAATLTERAKTISPGKNGYVVKVFAFGA
Qy 1197 VENPEYLTPOGGAPOHPHPAPAFSPAFDNLYYMDQDPERGAPSTPKGTLENPEYLG
Db 1191 VENPEYLTPOGGAPOHPHPAPAFSPAFDNLYYMDQDPERGAPSTPKGTLENPEYLG
Qy 1257 LDVEV 1261
Db 1251 LDVEV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
PN WO200153463-A2.
XX
PD 26-JUL-2001.
XX
PF 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
PA (CORI-) CORIXA CORP.
XX

```

```

PI Cheever MA, Hand-Zimmermann S;
XX
DR WPI; 2001-476112/51.
DR N-PSDB; AAB23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer.
XX
XX Claim 2; Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;
XX
Query Match 96.8%; Score 6621; DB 22; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;
Qy 1 MELALCRWGLLALLPQAASTVCTGTDMKLRPAPETHLDMRLHLYGCGVVGQNL 60
Db 1 MELALCRWGLLALLPQAASTVCTGTDMKLRPAPETHLDMRLHLYGCGVVGQNL 60
Qy 61 ELTYVPTNASSIFLQIDIEVQGVYLIANNOYOVLOSLRIYRGQLFEDNYALAVLNG 120
Db 61 ELTYVPTNASSIFLQIDIEVQGVYLIANNOYOVLOSLRIYRGQLFEDNYALAVLNG 120
Qy 121 DELNNTTPVTGASPGGLSELOLSLTELKSGVLIQRNPOLCYQDTILMKDIFHNQOLA 180
Db 121 DELNNTTPVTGASPGGLSELOLSLTELKSGVLIQRNPOLCYQDTILMKDIFHNQOLA 180
Qy 181 LTLIDNNSRACHPSPCKSGRCWSESEDCQSLTRIVCAGGCRKGRPLPTDCHQC 240
Db 181 LTLIDNNSRACHPSPCKSGRCWSESEDCQSLTRIVCAGGCRKGRPLPTDCHQC 240
Qy 241 AAGCTGPKXSDCLACLFHNSGICELHSPALVTYNTDFTESMPNPEGVYTGASCVTACP 300
Db 241 AAGCTGPKXSDCLACLFHNSGICELHSPALVTYNTDFTESMPNPEGVYTGASCVTACP 300
Qy 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSKPCARVCYGLGMQYIRANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSKPCARVCYGLGMQYIRANSKFIGIT 360
Qy 361 ELLEFAGCKKIFGSLAFLESPFGDPASNTAPLOPEQLQVFETLEETIGVLIYASAPDSLP 420
Db 361 IOEFAGCKKIFGSLAFLESPFGDPASNTAPLOPEQLQVFETLEETIGVLIYASAPDSLP 420
Qy 421 DLVSFQNTQVIRGRILHNGAVSLTLQIGISWLGRLSRLSGSLALIHNTLCEVATV 480
Db 421 DLVSFQNTQVIRGRILHNGAVSLTLQIGISWLGRLSRLSGSLALIHNTLCEVATV 480
Qy 481 PMDCLFMRPHQALLHTARPEDECVGBELACHQLCARHCHGPGTQCVNCSQFLRGDEC 540
Db 481 PMDCLFMRPHQALLHTARPEDECVGBELACHQLCARHCHGPGTQCVNCSQFLRGDEC 540
Qy 541 VEECEVLOGLPREYVNAASHCLPCHPECCPONGSVTCFPEADQCAVCAHYADPPECVARC 600
Db 541 VEECEVLOGLPREYVNAASHCLPCHPECCPONGSVTCFPEADQCAVCAHYADPPECVARC 600
Qy 601 PFNNFTVSFWLRVPVYASHE---PDEGACQCPPLNCTHSCVDLDDKGCPRAEQSPAS 656
Db 601 PSG-----VKPLSTMPIMKFPDEGACQPCPINCTHSCVDLDDKGCPRAEQSPAS 650
Qy 657 LTSIVSAVGIILVVVLGVFGILIKRROOKIRKTYMRLLQETELVEPLTPSGAMPNQA 716

```

Db 651 LTIISAVGILLVVGVGVGILIKRQKIRKYNRLQLQETELVEPLTPSGAMENQA 710  
 QY 717 QMRILKETELRKVVLGSGAFGTYYKGIWIPDGENVKIPAIKVLRENTSPKANKETLDE 776  
 Db 711 QMRILKETELRKVVLGSGAFGTYYKGIWIPDGENVKIPAIKVLRENTSPKANKETLDE 770  
 QY 777 AYVWAGVGSFVYSRLGICLTSTVQLVTLQMPYGCCLLDHVENRGRIGSODLLNMCQIA 836  
 Db 771 AYVWAGVGSFVYSRLGICLTSTVQLVTLQMPYGCCLLDHVENRGRIGSODLLNMCQIA 830  
 QY 837 KGMVLEDEVLRHDLAARVLYKSPVHKITDGLARLLDIDETETHAGGVPIKMA 896  
 Db 831 KGMVLEDEVLRHDLAARVLYKSPVHKITDGLARLLDIDETETHAGGVPIKMA 890  
 QY 897 LESILRRRFTHQSDVMSYGVTVLMTFGAKPYDGIIPAREIPDLEKGERLPQPICTID 956  
 Db 891 LESILRRRFTHQSDVMSYGVTVLMTFGAKPYDGIIPAREIPDLEKGERLPQPICTID 950  
 QY 957 VYIMVWCMIMIDECRRFRRELVSFSRVARDPQRFVYVIONEDLGPASPLDSTFYRSLLE 1016  
 Db 951 VYIMVWCMIMIDECRRFRRELVSFSRVARDPQRFVYVIONEDLGPASPLDSTFYRSLLE 1010  
 QY 1017 DDMGDLVDAEEYLVPOQGFCCDPAPAGAGMTHRRSSSTRSGGDLTLGSEPEEEA 1076  
 Db 1011 DDMGDLVDAEEYLVPOQGFCCDPAPAGAGMTHRRSSSTRSGGDLTLGSEPEEEA 1070  
 QY 1077 PRSPLABEGAGSDVDPDGLGMAKAGLQSLPHDPSLQRYSEDEPTVPLPSETDGYVAP 1136  
 Db 1071 PRSPLABEGAGSDVDPDGLGMAKAGLQSLPHDPSLQRYSEDEPTVPLPSETDGYVAP 1130  
 QY 1137 LTCSPOPEYVNOPDVRQPPSPREGPLPARPAGATLERAKTLSPKNGVYKVPFAGCA 1196  
 Db 1131 LTCSPOPEYVNOPDVRQPPSPREGPLPARPAGATLERAKTLSPKNGVYKVPFAGCA 1190  
 QY 1197 VENPEYLTPOGGAAPQHPAPSPAFDNLVYWDODPBERGAPSTFKGPTAENPEYLIG 1256  
 Db 1191 VENPEYLTPOGGAAPQHPAPSPAFDNLVYWDODPBERGAPSTFKGPTAENPEYLIG 1250  
 QY 1257 LDVPV 1261  
 Db 1251 LDVPV 1255  
 RESULT 10  
 AAG88267  
 ID AAG88267 standard; Protein; 1255 AA.  
 AC AAG88267;  
 DT 11-SEP-2001 (first entry)  
 DE HER2/neu amino acid sequence.  
 KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;  
 KW Immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;  
 KW Tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.  
 OS Homo sapiens.  
 PN MO200141787-A1.  
 PD 14-JUN-2001.  
 PF 11-DEC-2000; 2000MO-US33591.  
 PR 10-DEC-1999; 99US-0458299.  
 PA (EPIM-) EPIMONE INC.  
 PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E,  
 PI Keogh E;  
 XX MPI, 2001-374995/39.

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing  
 PT cellular immune responses for the prevention and treatment of cancer.  
 XX  
 PS Disclosure, Page 15; 1999p; English.

CC The present invention describes isolated prepared HER2/neu epitopes (I).  
 CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is  
 CC culture in vitro and binds to a complex of an epitope (II), bound to a  
 CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (1)  
 CC and a second epitope and the peptide is less than 50 contiguous amino  
 CC acids that have 100% identity with a native peptide sequence of HER2/neu;  
 CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical  
 CC excipient; (4) an isolated nucleic acid encoding a peptide comprising  
 CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic  
 CC and immunostimulant activities, and can be used in vaccines (I), (II)  
 CC and (III) are useful for inducing cellular immune responses for the  
 CC prevention and treatment of cancer. (I) and (II) are useful for  
 CC monitoring or evaluating an immune response to a tumour-associated  
 CC antigen when incubated with a T lymphocyte sample from a patient and  
 CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope  
 CC based vaccines mean that immunosuppressive epitopes that may be present  
 CC in whole antigens may be avoided. Selected epitopes may be combined to  
 CC enhance immunogenicity. The possible pathological side effects caused by  
 CC infectious agents or whole protein antigen is eliminated. The vaccine  
 CC provides the ability to direct and focus an immune response to multiple  
 CC selected antigens from the same pathogen. Epitope-based anti-tumour  
 CC vaccines provides the opportunity to combine epitopes derived from  
 CC multiple tumour-associated molecules to address the problem of tumour-  
 CC tumour variability and reducing the likelihood of tumour escape due to  
 CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in  
 CC the exemplification of the present invention.

Sequence 1255 AA;

Query Match 96.8%; Score 6621; DB 22; Length 1255;  
 Best Local Similarity 96.8%; Pred. No. 0;  
 Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

QY 1 MELALICRWGILLALPPGAASVCTCTDKMLRPLPASPETHLDMRLYLQCCQVYQCNL 60  
 Db 1 MELALICRWGILLALPPGAASVCTCTDKMLRPLPASPETHLDMRLYLQCCQVYQCNL 60  
 QY 61 ELTYLPNALSFLDIOEVGVYLIANQVQVPLQRLIRVGTQLEEDNYALAVLDNG 120  
 Db 61 ELTYLPNALSFLDIOEVGVYLIANQVQVPLQRLIRVGTQLEEDNYALAVLDNG 120  
 QY 121 DPLNNTFTVGTASREGARELQRLSTELIKGVLIQRPOLCYQDTIMKDIFFKNNOLA 180  
 Db 121 DPLNNTFTVGTASREGARELQRLSTELIKGVLIQRPOLCYQDTIMKDIFFKNNOLA 180  
 QY 181 LTLIDTNRSRACHPSPCKSRGWGESSSEDCSLTRTVACAGGACRGPPLPTCCHQC 240  
 Db 181 LTLIDTNRSRACHPSPCKSRGWGESSSEDCSLTRTVACAGGACRGPPLPTCCHQC 240  
 QY 241 AAGCTGPRGSHCLACLNHNSGICELRCPALVTYNTDTFESMPNPEGRTYTGASCVTACP 300  
 Db 241 AAGCTGPRGSHCLACLNHNSGICELRCPALVTYNTDTFESMPNPEGRTYTGASCVTACP 300  
 QY 301 YNYLSTDVGSCTVLCPLNCEVTAEDGTORCEKSKCARCYGLGMYIKANSKFIGIT 360  
 Db 301 YNYLSTDVGSCTVLCPLNCEVTAEDGTORCEKSKCARCYGLGMYIKANSKFIGIT 360  
 QY 361 ELRFAGCKKIFGSLAFIPESFDGDPASNTABLOEOLQVETLEITGYLYISAMPDLP 420  
 Db 361 IQEFAGCKKIFGSLAFIPESFDGDPASNTABLOEOLQVETLEITGYLYISAMPDLP 420  
 QY 421 DLSVFQNLQVIRGRILNNGASLTLOGLIGISWGLRSLRELGSGLAIHNHNLCPVHTV 480  
 Db 421 DLSVFQNLQVIRGRILNNGASLTLOGLIGISWGLRSLRELGSGLAIHNHNLCPVHTV 480  
 QY 481 PMDQLFENPHOALHTANREDECVGGLACHQLCARGHGCMGPPTOCVNCOSPLRGOEC 540  
 Db 481 PMDQLFENPHOALHTANREDECVGGLACHQLCARGHGCMGPPTOCVNCOSPLRGOEC 540

```

Db      481 PMDLPRNHQALLHTANRPEDECVBGLAChOLCARGHCMGPGPTQCVCNCSQFLRGQEC 540
Qy      541 VEECRVILQGLPREVYNARHCLPCHPECCQPNQSVTCFPGPADCCVACAHYKDPFCVABC 600
Db      541 VEECRVILQGLPREVYNARHCLPCHPECCQPNQSVTCFPGPADCCVACAHYKDPFCVABC 600
Qy      601 PFNFNTVCSFWLVRPKVKSASHLE-----PDEGACQPCPINTHSCVLDLDDGCAEGRASP 656
Db      601 PSC-----VKPRLSTMPIMKFPDEGACQPCPINTHSCVLDLDDGCAEGRASP 650
Qy      657 LSTISAVNGILLVVLGVVFGILIKRQCKIRKRYTRRLLOETEVEPLTPSGAMPNGA 716
Db      651 LSTISAVNGILLVVLGVVFGILIKRQCKIRKRYTRRLLOETEVEPLTPSGAMPNGA 710
Qy      717 QMRILKTELKRVKVGSGAFGTVYKGIWIPDGENVKIPALIVLENSPKANKILDE 776
Db      711 QMRILKTELKRVKVGSGAFGTVYKGIWIPDGENVKIPALIVLENSPKANKILDE 770
Qy      777 AYVWAGVSPVSRLLGICLTSTVQLTQMLPYGCLLDHYENRGRGLSGQDLNMCQIA 836
Db      771 AYVWAGVSPVSRLLGICLTSTVQLTQMLPYGCLLDHYENRGRGLSGQDLNMCQIA 830
Qy      837 KGSMSYLEDVRLVHRDLAARVLYKSPNHVKITPGLARLIDETETVHADGKVPIKMA 896
Db      831 KGSMSYLEDVRLVHRDLAARVLYKSPNHVKITPGLARLIDETETVHADGKVPIKMA 890
Qy      897 LESTLRRTFHOSDWSYGVTVWELMTFGAKPYDGIPIREIPLLEKGERLPQPICTID 956
Db      891 LESTLRRTFHOSDWSYGVTVWELMTFGAKPYDGIPIREIPLLEKGERLPQPICTID 950
Qy      957 VYIMIVKCMWIDSECRFRELVESESRMARDPQRFVIVIONEDLGASPLDSTFYRSLE 1016
Db      951 VYIMIVKCMWIDSECRFRELVESESRMARDPQRFVIVIONEDLGASPLDSTFYRSLE 1010
Qy      1017 DDMGDLVDAEEXYVPOGGFPCPPAPAGAGMHHRRSSSTSGGDLTLGEPSEEA 1076
Db      1011 DDMGDLVDAEEXYVPOGGFPCPPAPAGAGMHHRRSSSTSGGDLTLGEPSEEA 1070
Qy      1077 PRSLPASEGASGVDFDGLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSESTDYVAP 1136
Db      1071 PRSLPASEGASGVDFDGLGMAKAGLQSLPTHDPSPLORYSEDPVPLPSESTDYVAP 1130
Qy      1137 LTGSPQPEYVNPQVPRQPPSPREGPLPAARPAAGATLERAKTISPGKNGVYKQVAPFGA 1196
Db      1131 LTGSPQPEYVNPQVPRQPPSPREGPLPAARPAAGATLERAKTISPGKNGVYKQVAPFGA 1190
Qy      1197 VENPEXYLTPOGGAAPQHPAPAFADNLVYWDQDPERGAPSTPKGTPTANDEYILG 1256
Db      1191 VENPEXYLTPOGGAAPQHPAPAFADNLVYWDQDPERGAPSTPKGTPTANDEYILG 1250
Qy      1257 LDVAV 1261
Db      1251 LDVAV 1255

```

## RESULT 11

AAE24067 standard; Protein; 1255 AA.

AAE24067;

23-SEP-2002 (first entry)

Human Her-2 protein.

Human Her-2; epidermal growth factor receptor 2; infection; cancer;

hyperproliferative disorder; prophylaxis; inflammation; antisense;

tumour; gene therapy; phosphorothioate backbone.

Homo sapiens.

W0200222636-A1.

```

PD      21-MAR-2002.
XX      12-SEP-2001; 2001WO-0S28572.
XX      15-SEP-2000; 2000US-0663834.
XX      (ISIS-) ISIS PHARM INC.
PA      Bennett CF, Coweart LM;
XX      WPI; 2002-471192/50.
XX      N-PSDB; AAD38904.
PT      Novel antisense oligonucleotide which modulates the expression of Human
PT      Epidermal Growth Factor receptor, Her2, is useful for treating tumors
PT      inflammation or to prevent infection in humans.
XX      Example 13; Page 95-107; 116pp; English.

```

The invention relates to antisense compounds targeted to a nucleic acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2) that specifically hybridises with and inhibits the expression of Her2. Antisense compounds of the invention are used for treating diseases or conditions associated with Her2 such as hyperproliferative disorders e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary, CC neural or cardiac cancer. They are also useful prophylactically e.g. to prevent or delay infection, inflammation and tumour formation. The invention is also used in gene therapy. The present sequence is human Her-2 protein.

Sequence 1255 AA;

Query Match 96.8%; Score 6621; DB 23; Length 1255;

Best Local Similarity 96.8%; Pred. No. 0; Mismatches 19; Indels 14; Gaps 2;

```

Qy      1 MELALCRWGLLALLPQAASTVCTGTMKRLPASPETHLDRLHYQCCVYQGNL 60
Db      1 MELALCRWGLLALLPQAASTVCTGTMKRLPASPETHLDRLHYQCCVYQGNL 60
Qy      61 ELTYPTNALSFLDDIOEVQGVYLIANNQVQVLOQLRIYRGQLFEDNYALAVLNG 120
Db      61 ELTYPTNALSFLDDIOEVQGVYLIANNQVQVLOQLRIYRGQLFEDNYALAVLNG 120
Qy      121 DELNNTPTAGSPGGLRELOLRSLTEILKGVLIQRNPOLCYCTIIMKDIFFKNNOLA 180
Db      121 DELNNTPTAGSPGGLRELOLRSLTEILKGVLIQRNPOLCYCTIIMKDIFFKNNOLA 180
Qy      181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLETDCHEQC 240
Db      181 LTLIDNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVCAAGCARCKGPLETDCHEQC 240
Qy      241 AAGCTGPRHSDCLACTHNHSGICELHCPALVTYNTDTFESMPNPEGYTGCASVTACP 300
Db      241 AAGCTGPRHSDCLACTHNHSGICELHCPALVTYNTDTFESMPNPEGYTGCASVTACP 300
Qy      301 YNYLSTDVGSCTLVCPILHNOEVTAEAGTORCEKSKPCARVYGLAMQYIRANSKFISIT 360
Db      301 YNYLSTDVGSCTLVCPILHNOEVTAEAGTORCEKSKPCARVYGLAMQYIRANSKFISIT 360
Qy      361 ELLEPAGCKKIFGSLAFLEPSTFGDPASTATPLOPQLOVFEETLGTGLYISAMPDLP 420
Db      361 IDEFAGCKKIFGSLAFLEPSTFGDPASTATPLOPQLOVFEETLGTGLYISAMPDLP 420
Qy      421 DLVSFQNLQVIRGRIIHNGAYSILTQGGISMLGLRSRLREGSGLATLHNHTHLCFYVTV 480
Db      421 DLVSFQNLQVIRGRIIHNGAYSILTQGGISMLGLRSRLREGSGLATLHNHTHLCFYVTV 480
Qy      481 FMDQLFRNPHQALLHTANRPEDECVBGLAChOLCARGHCMGPGPTQCVCNCSQFLRGQEC 540
Db      481 FMDQLFRNPHQALLHTANRPEDECVBGLAChOLCARGHCMGPGPTQCVCNCSQFLRGQEC 540
Qy      541 VEECRVILQGLPREVYNARHCLPCHPECCQPNQSVTCFPGPADCCVACAHYKDPFCVABC 600

```

```

Db      541 VEECRVLTQGLPREYVNAHCLPCHPECPQNSGVTFCFBEADQCVACAHYKDPFCVAC 600
Qy      601 PENNFVSMFLMRPKYSASHLE-----PDEEGACOPCPINCTSCVDLDKGPABEASAP 656
Db      601 PSC-----VKPDLSTMPWKPFDEBGAQCPINCTSCVDLDKGPABEASAP 650
Qy      657 LFSIYSAVVGILLVVLGVFGILLIKRQOKIRKYMTRRLLOETLEVELETPSGAMPNOA 716
Db      651 LFSIISAVVGILLVVLGVFGILLIKRQOKIRKYMTRRLLOETLEVELETPSGAMPNOA 710
Qy      717 QMRILKETELRKXKVLGSAFGVYKGIIPDGENKIVALKVLAENISPAKNEILDE 776
Db      711 QMRILKETELRKXKVLGSAFGVYKGIIPDGENKIVALKVLAENISPAKNEILDE 770
Qy      777 AAYMAGVSPYVSRLLGICLTSTVQLTOLMPYGCILDHVRENGRLSGQDILLNMCQIA 836
Db      771 AAYMAGVSPYVSRLLGICLTSTVQLTOLMPYGCILDHVRENGRLSGQDILLNMCQIA 830
Qy      837 KGMSTYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPDKMA 896
Db      831 KGMSTYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLLDIDETRYHADGKVPDKMA 890
Qy      897 LESILRRFTHSDVMSYGVYWEIMTFGAKPYDGIIPAREIDPLEKGRLEQPPICITD 956
Db      891 LESILRRFTHSDVMSYGVYWEIMTFGAKPYDGIIPAREIDPLEKGRLEQPPICITD 950
Qy      957 VYIMVWKCMIDSECPREFRELVSFSRMAQDPQRFVVIQNEIDLGPASPLDSTFYRLLE 1016
Db      951 VYIMVWKCMIDSECPREFRELVSFSRMAQDPQRFVVIQNEIDLGPASPLDSTFYRLLE 1010
Qy      1017 DDDMGDLVAEELVAVQGGFPCDPAPAGAGVHHRRSSSTRSGGDTLTGLSESEEA 1076
Db      1011 DDDMGDLVAEELVAVQGGFPCDPAPAGAGVHHRRSSSTRSGGDTLTGLSESEEA 1070
Qy      1077 PRSPAPSEAGSDVDFDGLGNAKAGQSLPTHPSPLOQRYSEDPTVPLPSETDGYAP 1136
Db      1071 PRSPAPSEAGSDVDFDGLGNAKAGQSLPTHPSPLOQRYSEDPTVPLPSETDGYAP 1130
Qy      1137 LTCSPQPEYVNPDPVAPQPPSPREGPLPAAPAGATLEBAKTLSPGKGVAVDPAFGCA 1196
Db      1131 LTCSPQPEYVNPDPVAPQPPSPREGPLPAAPAGATLEBAKTLSPGKGVAVDPAFGCA 1190
Qy      1197 VENPEYLTPQGAAPQPPPAFPAFDFLYWDDPPRGAQPSPTFGKPTPAENPEYLIG 1256
Db      1191 VENPEYLTPQGAAPQPPPAFPAFDFLYWDDPPRGAQPSPTFGKPTPAENPEYLIG 1250
Qy      1257 LDVPV 1261
Db      1251 LDVPV 1255

```

RESULT 12  
AAE20479 standard; Protein, 1255 AA.

```

ID      AAE20479 standard; Protein, 1255 AA.
XX      AAE20479;
AC      01-JUL-2002 (first entry)
DT      01-JUL-2002 (first entry)
XX      Human Her-2/neu protein.
DE      Human Her-2/neu protein.
KW      Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
OS      human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX      Homo sapiens.
FH      Key
FT      Location/Qualifiers
XX      1021..1030
XX      /note="Naturally processed HLA-B44-restricted epitope"
XX      WO200214503-A2.

```

```

PD      21-FEB-2002.
XX      14-AUG-2001; 2001WO-US41733.
PF      14-AUG-2001; 2000US-225152P.
XX      14-AUG-2000; 2000US-225152P.
PR      28-SEP-2000; 2000US-236428P.
XX      21-FEB-2001; 2001US-270520P.
XX      (CORI-) CORIXA CORP.
PI      Hand-zimmermann S, Cheever MA, Foy TM, Lodes NJ, Kalos MD;
XX      McNeill PD, Vedyck TS;
XX      WPI; 2002-280758/32.
XX      N-PSDB; AAD32743.
XX      Novel isolated Her-2/Neu polypeptide composition useful for therapy,
XX      prevention and diagnosis of cancer, preferably breast cancer -
XX      Disclosure; Page 114-117; 129pp; English.
XX      The invention relates to an isolated Her-2/Neu polypeptide composition
XX      effective for eliciting an immune response. The invention is useful for
XX      eliciting an immune response in a patient, where the patient is human
XX      leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX      The composition is useful for the therapy and diagnosis of cancer,
XX      preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX      and other compositions for the diagnosis, prevention and treatment of
XX      human malignancies, for stimulating and/or expanding T cells specific for
XX      Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX      patient. The invention is useful for stimulating a T cell response in a
XX      human patient, as probe or primer for nucleic acid hybridisation, to
XX      selectively form duplex molecules with complementary stretches of the
XX      entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX      length gene from a suitable library, and to direct expression of a
XX      polypeptide in appropriate host cells. The composition is useful in
XX      prophylactic or therapeutic applications and for the treatment of cancer,
XX      preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX      associated malignancies. The invention is useful in gene therapy. The
XX      present sequence is human Her-2/neu protein.
SQ      Sequence 1255 AA;
Query Match 96.8%; Score 6621; DB 23; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;
Qy      1 MELAALCRMGILLALLPFGAASVQVCTGTMKRLPASPETHLDMRLHYGCGVQGNL 60
Db      1 MELAALCRMGILLALLPFGAASVQVCTGTMKRLPASPETHLDMRLHYGCGVQGNL 60
Qy      61 ELTYLPTNASSLSFLQDIQEVQGYVLAHNOVQVPLQRLIRVGTQJFEDNYALAVLDNG 120
Db      61 ELTYLPTNASSLSFLQDIQEVQGYVLAHNOVQVPLQRLIRVGTQJFEDNYALAVLDNG 120
Qy      121 DPLNNTPTVAGSPGGRRELQSLSTETILKGVYLQNPOLCYQDTLLMDIFPKKNQLA 180
Db      121 DPLNNTPTVAGSPGGRRELQSLSTETILKGVYLQNPOLCYQDTLLMDIFPKKNQLA 180
Qy      181 LTLIDTNRSPACHPCSPCKGSRGWSSESDCSLRTVAGGACGKGLPTDCHEQC 240
Db      181 LTLIDTNRSPACHPCSPCKGSRGWSSESDCSLRTVAGGACGKGLPTDCHEQC 240
Qy      241 AAGCTGKMSDCLACHFNHSGICEHLCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Db      241 AAGCTGKMSDCLACHFNHSGICEHLCPALVYNTDTFESMPNPEGRYTFGASCVTACP 300
Qy      301 YNYLSTVGSCTLVCPAHNOEYTAEDTQCEKSCPKCAVCYGLGQVYKANSKFIGIT 360
Db      301 YNYLSTVGSCTLVCPAHNOEYTAEDTQCEKSCPKCAVCYGLGQVYKANSKFIGIT 360
Qy      361 ELEPAGCKIFGSLAPLPESFGDPAASNTAPLOPEQLQVETLEETITGYLYISAMPDGLP 420

```



```

Db      361 IOEFACCKIFSGSLAFESFDGSDPASNTAPLQPEOLQVFTLEETITGLYISAMPDSL 420
Qy      421 DLSVFONLQVIRGRIILHNGAYSITLQGLISWLGRLSELGSGALIHNNHLCFVHTV 480
Db      421 DLSVFONLQVIRGRIILHNGAYSITLQGLISWLGRLSELGSGALIHNNHLCFVHTV 480
Qy      481 PMDQLFRNPQALHTANRPEDECVGEGALCHQLCARQWQGGPQCVCNCSQFLRGQSC 540
Db      481 PMDQLFRNPQALHTANRPEDECVGEGALCHQLCARQWQGGPQCVCNCSQFLRGQSC 540
Qy      541 VEECRVLOGLPREYVNAHCLPCHPECPQNSVTCFGEADQVCAHAHYKDPFCVAC 600
Db      541 VEECRVLOGLPREYVNAHCLPCHPECPQNSVTCFGEADQVCAHAHYKDPFCVAC 600
Qy      601 PFNNFTVSEFLRKPASASLE----PDEGACOPICICTSCVLDLXKGPABRASP 656
Db      601 PSC-----VKPDLSTYPIWKFPDEGACQCPINCTSCVLDLXKGPABRASP 650
Qy      657 LRSIVSAVGLLVVVLGVVFGILIRKQOKIRKYMRLLOETELVEBLTSGAMPNQA 716
Db      651 LRSIVSAVGLLVVVLGVVFGILIRKQOKIRKYMRLLOETELVEBLTSGAMPNQA 710
Qy      717 QMRILKETELRKKTUGSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPANKIIDE 776
Db      711 QMRILKETELRKKTUGSGAFGTVYKGIWIPDGENVKIPVALKVLRENTSPANKIIDE 770
Qy      777 AYMAGVSPYVSRLLGICLTSTVQLTQMPYGLLDHVENRGLSGODLLNMCQJIA 836
Db      771 AYMAGVSPYVSRLLGICLTSTVQLTQMPYGLLDHVENRGLSGODLLNMCQJIA 830
Qy      837 KGSYLEDVRLVHRDLAARNVLKSPNHKITYDFGLRLIDIDETIYHADGGKVPKXMA 896
Db      831 KGSYLEDVRLVHRDLAARNVLKSPNHKITYDFGLRLIDIDETIYHADGGKVPKXMA 890
Qy      897 LESILRRFTHQSDVMSYGVTVMLTFFGAKPYDGI PAEIPDLLEKGERLEPOPICTID 956
Db      891 LESILRRFTHQSDVMSYGVTVMLTFFGAKPYDGI PAEIPDLLEKGERLEPOPICTID 950
Qy      957 VYIMWKCMWIDSECRPRELIVSEFSRMAADPQRFVYIQNEGLGASPLDSTFYSLLE 1016
Db      951 VYIMWKCMWIDSECRPRELIVSEFSRMAADPQRFVYIQNEGLGASPLDSTFYSLLE 1010
Qy      1017 DDMGDLVAEEYLVQGGFCFCDPAFGAGMWHHRSSSTSGGGDLTLGIESEEA 1076
Db      1011 DDMGDLVAEEYLVQGGFCFCDPAFGAGMWHHRSSSTSGGGDLTLGIESEEA 1070
Qy      1077 PRSPLAPSEGAGSDVFDGLGMAAGLQSLPTHDESPLORYSEDPVLPSEDTGYVAP 1136
Db      1071 PRSPLAPSEGAGSDVFDGLGMAAGLQSLPTHDESPLORYSEDPVLPSEDTGYVAP 1130
Qy      1137 LITSPQBEYVNOVDVPPSPREGPLPARPAGATLERAKTISFGKGVYKDVAFGGA 1196
Db      1131 LITSPQBEYVNOVDVPPSPREGPLPARPAGATLERAKTISFGKGVYKDVAFGGA 1190
Qy      1197 VENPEYLTPOGGAAPQHPAPSPAFDNLVWDQPPERGAAPSTFKGPTANENYLG 1256
Db      1191 VENPEYLTPOGGAAPQHPAPSPAFDNLVWDQPPERGAAPSTFKGPTANENYLG 1250
Qy      1257 LDVPV 1261
Db      1251 LDVPV 1255

```

RESULT 13  
AAMS1143  
ID AAMS1143 standard; Protein; 1255 AA.

AC AAM1143;  
XX  
DT 17-JUN-2002 (first entry)  
XX  
DE Human Her-2/neu oncogene-encoded p185 glycoprotein.

```

KW Her-2/neu: oncogene; cancer; tumour; vaccine; human; p185;
KM tyrosine kinase; receptor; c-erbB2; gene therapy.
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT Domain 1..653
FT FT /note="extracellular domain"
FT FT 676..1255
FT FT /note="intracellular domain"
FT FT 990..1255
FT FT /note="phosphorylation domain"
PN Domain
PN WO200212341-A2.
PD 14-FEB-2002.
PF 03-AUG-2001; 2001WO-US24283.
PF 03-AUG-2000; 2000US-0632507.
PA (CORI-) CORIYA CORP.
PA (SMK) SMITHKLINE BEECHAM BIOLOGICALS.
PI Cheever MA, Gheysen D;
PI WPI, 2002-241743/29.
PI N-PESDB; ABA92250.
PT Her-2/neu fusion protein for treating or preventing cancer by eliciting
PT or enhancing an immune response to the protein, has Her-2/neu
PT extracellular domain fused to Her-2/neu intracellular or
PT phosphorylation domain
PS Claim 68; Fig 7; 141pp; English.
XX
XX The present sequence is that of human Her-2/neu (p185 glycoprotein
XX or c-erbB2), an oncogenic self-protein and target for anti-cancer
XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
XX in a variety of cancers, including breast, ovarian, colon, lung and
XX prostate cancer. Her-2/neu is a member of the tyrosine kinase
XX family of receptor-like glycoproteins. It comprises an extracellular
XX domain with homology to the epidermal growth factor receptor
XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
XX intracellular domain that also shows homology to EGFR. Its
XX overexpression correlates with a poor prognosis in breast and
XX ovarian cancers. The invention provides Her-2/neu fusion
XX proteins, nucleic acids encoding them, viral vectors, and vaccines
XX comprising the fusion proteins or nucleic acid molecules. In
XX preferred fusion proteins, the extracellular domain of a Her-2/neu
XX protein is fused to a Her-2/neu intracellular domain or
XX phosphorylation domain (or its DeltaCD fragment). An immune
XX response to Her-2/neu protein is elicited or enhanced by
XX transfecting the fusion protein in the form of a vaccine, or by
XX transfecting cells of an animal ex vivo with a nucleic acid
XX encoding the fusion protein, and delivering the transfected cells
XX to the animal. The fusion proteins, nucleic acids, and isolated
XX specific T-cells are useful for inhibiting the development of a
XX cancer, especially breast, ovarian, colon, lung or prostate cancer
XX in a patient. T cells that specifically react with a Her-2/neu
XX fusion protein can be used to remove tumour cells from a sample in
XX order to inhibit the development of cancer in a patient.
XX
XX Sequence 1255 AA;

```

Query Match 96.8%; Score 6621; DB 23; Length 1255;  
Best Local Similarity 96.8%; Pred. No. 0;  
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

Qy 1 MELALACRWGLLLLPFGASTGVCTGTDMKRLPASPETHLMKRLHYGCGVQGNL 60  
Db 1 MELALACRWGLLLLPFGASTGVCTGTDMKRLPASPETHLMKRLHYGCGVQGNL 60

```

QY 61 ELTYLPTNASLSLQDIOEVQGYVLLIAHNVQROVPLQRLIRIVGTQLFEDNVALAVLDNG 120
Db 61 ELTYLPTNASLSLQDIOEVQGYVLLIAHNVQROVPLQRLIRIVGTQLFEDNVALAVLDNG 120
QY 121 DPLNNTTPVTGASPGGLRELOQLASTLEIIKGGVLIQNPOLCYQDPTILMKDIFHKNOQA 180
Db 121 DPLNNTTPVTGASPGGLRELOQLASTLEIIKGGVLIQNPOLCYQDPTILMKDIFHKNOQA 180
QY 181 LTLIDNRSRACHPCSPCKGSRCWGESSEDCOSLTRVTCAGACGACGKPLPTDCHEQC 240
Db 181 LTLIDNRSRACHPCSPCKGSRCWGESSEDCOSLTRVTCAGACGACGKPLPTDCHEQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTRESMNPREGRTYFGASCTYACP 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTRESMNPREGRTYFGASCTYACP 300
QY 301 YNYLSTDVGSCTLVCPHNOEVAEDGTQCEKSPCARVCGLGMOYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHNOEVAEDGTQCEKSPCARVCGLGMOYIKANSKFIGIT 360
QY 361 ELFFAGCKTIFGSLAFLPESFDGDPASNTAPLOPELOVETLEEITGYLYTSAMPDULP 420
Db 361 IOEFAGCKTIFGSLAFLPESFDGDPASNTAPLOPELOVETLEEITGYLYTSAMPDULP 420
QY 421 DLSVFONLOYIRGRILHNGAYSLTLQGLISWLGASLEIGSGALIHNTHLCTVHTV 480
Db 421 DLSVFONLOYIRGRILHNGAYSLTLQGLISWLGASLEIGSGALIHNTHLCTVHTV 480
QY 481 PMDOLFRRPHOALHTANRPEDECVGEGLACHOLCARGHGWGPGPTQVCNCSQFLRGQEC 540
Db 481 PMDOLFRRPHOALHTANRPEDECVGEGLACHOLCARGHGWGPGPTQVCNCSQFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAARHCLPCHPECOQNGSVTFGEBAOCVACAHYKXDPFVAVAC 600
Db 541 VEECRVLOGLPREYVNAARHCLPCHPECOQNGSVTFGEBAOCVACAHYKXDPFVAVAC 600
QY 601 PENNFVSFWLRYPKVYASHL-----PDEGACQPPINCTHSVDLDDKGCAPABRASP 656
Db 601 PEG-----VKRDLSYMPIWKFPDEGACQPPINCTHSVDLDDKGCAPABRASP 650
QY 657 LNSISAVVGLLVVVGAVFGILLIRROQKIRKYMRLLQETLVEBPLPSGAMPNQA 716
Db 651 LNSISAVVGLLVVVGAVFGILLIRROQKIRKYMRLLQETLVEBPLPSGAMPNQA 710
QY 717 QMRILKETELRKVYVLSGAFGVYVGIWIPGENVKIYAIKVLRENSPANKELIDE 776
Db 711 QMRILKETELRKVYVLSGAFGVYVGIWIPGENVKIYAIKVLRENSPANKELIDE 770
QY 777 AYVMAVGSPYVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSODLLNMCQIA 836
Db 771 AYVMAVGSPYVSRLLGICLTSTVQLVTLQMPYGLLDHVRNRRGLSODLLNMCQIA 830
QY 837 KGMSTYEDVRLVHRDLAARNVLKSPNHKIDTDFGLARLLDDEIFYHADGKRVIKMA 896
Db 831 KGMSTYEDVRLVHRDLAARNVLKSPNHKIDTDFGLARLLDDEIFYHADGKRVIKMA 890
QY 897 LESILRRRTTHQSDVWSYGVTWELMTFGAKPYDGIPIASEIPDLLEKGRPLPOPICITD 956
Db 891 LESILRRRTTHQSDVWSYGVTWELMTFGAKPYDGIPIASEIPDLLEKGRPLPOPICITD 950
QY 957 VYMWIKCMIMISEGPRPRELVESESRMARPPORFVITQNEDELCPASPLDSTFYRSILE 1016
Db 951 VYMWIKCMIMISEGPRPRELVESESRMARPPORFVITQNEDELCPASPLDSTFYRSILE 1010
QY 1017 DDDMGDLVDAEYLYPQOGFPDPAAPGAGVHRHRSSTRSGGSLTLGLEPSEBEA 1076
Db 1011 DDDMGDLVDAEYLYPQOGFPDPAAPGAGVHRHRSSTRSGGSLTLGLEPSEBEA 1070
QY 1077 PRSPLAPSEGASDVFDGDLGGAAGKIQSLPETHDPSPLQRYSEDPTVPLPSETGYVAP 1136
Db 1071 PRSPLAPSEGASDVFDGDLGGAAGKIQSLPETHDPSPLQRYSEDPTVPLPSETGYVAP 1130
QY 1137 LTCSPQPEYVNPQDVRPQPPRBPGLPAARPAATLERAKTILSPGKNGVYKDVAFAGGA 1196

```

```

Db 1131 LTCSPQPEYVNPQDVRPQPPRBPGLPAARPAATLERAKTILSPGKNGVYKDVAFAGGA 1190
QY 1197 VENPEYLLTPQGGAAQPHPPAFSAFNLTYWDDPPERKGAPESTFGTPTAAKPEYLG 1256
Db 1191 VENPEYLLTPQGGAAQPHPPAFSAFNLTYWDDPPERKGAPESTFGTPTAAKPEYLG 1250
QY 1257 LDVPV 1261
Db 1251 LDVPV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein; 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001WO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
XX
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S;
XX
DR MPI; 2002-280741/32.
XX
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide.
XX
PS Disclosure; Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC comprising an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;

Query Match 96.8%; Score 6621; DB 23; Length 1255;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1225; Conservative 7; Mismatches 19; Indels 14; Gaps 2;

QY 1 MELAAALCRWGLLALPBGASQVCTGTMKRLPASPTHTDMLRHLYOGGQVYQGNL 60
Db 1 MELAAALCRWGLLALPBGASQVCTGTMKRLPASPTHTDMLRHLYOGGQVYQGNL 60
QY 61 ELTYLPTNASLSLQDIOEVQGYVLLIAHNVQROVPLQRLIRIVGTQLFEDNVALAVLDNG 120

```

```

Db      61 ELTYLPTNASTSLQIOEVQYVLI.AHNQVRQVPLQRIRVGTQLFEDNYALAVLDNG 120
Qy      121 DPLNNTPTVTGASPGGARELOASTLSTELIKGAVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
Db      121 DPLNNTPTVTGASPGGARELOASTLSTELIKGAVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
Qy      181 LTLIDNRSRACHPCSPMCKGSRGWGSESSDCQSLTRTVCAAGCARGKPLPTDCHEQC 240
Db      181 LTLIDNRSRACHPCSPMCKGSRGWGSESSDCQSLTRTVCAAGCARGKPLPTDCHEQC 240
Qy      241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTESMNPFGRTTFCGASCTYACP 300
Db      241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDTESMNPFGRTTFCGASCTYACP 300
Qy      301 YNYLSTDVGSCTLVCPPLNQEVTAEQDTCRCKSPCARVCYGLMGOYIKANSKEFIGIT 360
Db      301 YNYLSTDVGSCTLVCPPLNQEVTAEQDTCRCKSPCARVCYGLMGOYIKANSKEFIGIT 360
Qy      361 ELFFAGCKTIFGSLAF.PESFDDPASNTPALOPELOVFEETLEETGYLYISAMPDILP 420
Db      361 ELFFAGCKTIFGSLAF.PESFDDPASNTPALOPELOVFEETLEETGYLYISAMPDILP 420
Qy      421 DLSVFONTQYIRGRILHNGAYSLTLOGLISWLGRLSRLSGSLALIHNNTHLCFVHTV 480
Db      421 DLSVFONTQYIRGRILHNGAYSLTLOGLISWLGRLSRLSGSLALIHNNTHLCFVHTV 480
Qy      481 PMDOLFRNPHQALHTNRPEDCEVGBGLACHOLCARGHGWGPGPOCVNCSOPLRGQEC 540
Db      481 PMDOLFRNPHQALHTNRPEDCEVGBGLACHOLCARGHGWGPGPOCVNCSOPLRGQEC 540
Qy      541 VEECRVLOGLPREVYNARHCLPCHPEQOPNGSVTCFGEADOCVACAHYKDPFCVARG 600
Db      541 VEECRVLOGLPREVYNARHCLPCHPEQOPNGSVTCFGEADOCVACAHYKDPFCVARG 600
Qy      601 PENNFVSWLRVPCVSAHLE---PDEGACQPCPINCTHSCVDLDKQCPAEGRASP 656
Db      601 PSC-----VKSDLSYMPIMKPFDEGACQPCPINCTHSCVDLDKQCPAEGRASP 650
Qy      657 LTSIVAVVGIILVWVGVFGILIRKROOKIRKYMRLLOSTELVEPLTPSGAMPDA 716
Db      651 LTIISAVVGIILVWVGVFGILIRKROOKIRKYMRLLOSTELVEPLTPSGAMPDA 710
Qy      717 QMRILKTELKRYKVLGSAFGTVYKGIWIPDENYKIPVAIYKRENTSPKANKELIDE 776
Db      711 QMRILKTELKRYKVLGSAFGTVYKGIWIPDENYKIPVAIYKRENTSPKANKELIDE 770
Qy      777 AYVMAGVSPYVSRLLGICLTSTVOLVTQLMFYGCLLDHRENRGLSGSODLNMCMQIA 836
Db      771 AYVMAGVSPYVSRLLGICLTSTVOLVTQLMFYGCLLDHRENRGLSGSODLNMCMQIA 830
Qy      837 KGMSTYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLDIDETEYHADGKVPYIKMMA 896
Db      831 KGMSTYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLDIDETEYHADGKVPYIKMMA 890
Qy      897 LESILRRRFTHQSDWSYGTVWELMTFGAKPYDGIIPAREIPLLLEKGERLPPPTCTID 956
Db      891 LESILRRRFTHQSDWSYGTVWELMTFGAKPYDGIIPAREIPLLLEKGERLPPPTCTID 950
Qy      957 VVMIMVCMWIDSECRFPRELVESEFSRMARDPQRFVIONEDLGPASPIDSTFYSLLE 1016
Db      951 VVMIMVCMWIDSECRFPRELVESEFSRMARDPQRFVIONEDLGPASPIDSTFYSLLE 1010
Qy      1017 DDDMGDLVDAEYIYVPOQGFCDPAPAGAGMYHHRRSSSTSGGDDLTLGLEBSEBEA 1076
Db      1011 DDDMGDLVDAEYIYVPOQGFCDPAPAGAGMYHHRRSSSTSGGDDLTLGLEBSEBEA 1070
Qy      1077 PRSPLAPSEAGSDVFPDGLGMMGAAKGLQSLPTHDSPLQRYSESDPTVPLPSESTDYVAP 1136
Db      1071 PRSPLAPSEAGSDVFPDGLGMMGAAKGLQSLPTHDSPLQRYSESDPTVPLPSESTDYVAP 1130
Qy      1137 LTGSPQPEYVNOQDVRPQPSPREGPIPAARPAATLIERAKTILSPKNGVYKQVFAFGA 1196

```

```

Db      1131 LTGSPQPEYVNOQDVRPQPSPREGPIPAARPAATLIERAKTILSPKNGVYKQVFAFGA 1190
Qy      1197 VENEPEYLPQGAAPQPHPPAFSPAFDNLVYMDQDPEERGAPSTKGTFTANPEYL 1256
Db      1191 VENEPEYLPQGAAPQPHPPAFSPAFDNLVYMDQDPEERGAPSTKGTFTANPEYL 1250
Qy      1257 LDVPPY 1261
Db      1251 LDVPPY 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein, 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN MO9316185-A.
XX
PI 19-AUG-1993.
XX
PF 05-FEB-1993; 93WO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CEU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston TX, Huston JS, Oppermann H, Ring DB,
XX
DR MPI; 1993-272889/34.
XX
DR N-PSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
XX
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 96.2%; Score 6578; DB 14; Length 1433;
Best Local Similarity 96.3%; Fred. No. 0;
Matches 1218; Conservative 9; Mismatches 24; Indels 14; Gaps 2;

Qy      1 MELALCGMGLLALLPPGAASVCTGTDKRLPASPETHLMLRLVYQGVQGNL 60
Db      1 MELALCGMGLLALLPPGAASVCTGTDKRLPASPETHLMLRLVYQGVQGNL 60
Qy      61 ELTYLPTNASTSLQIOEVQYVLI.AHNQVRQVPLQRIRVGTQLFEDNYALAVLDNG 120
Db      61 ELTYLPTNASTSLQIOEVQYVLI.AHNQVRQVPLQRIRVGTQLFEDNYALAVLDNG 120
Qy      121 DPLNNTPTVTGASPGGARELOASTLSTELIKGAVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
Db      121 DPLNNTPTVTGASPGGARELOASTLSTELIKGAVLIQRNPOLCYQDTILMKDIFHKNNOLA 180
Qy      181 LTLIDNRSRACHPCSPMCKGSRGWGSESSDCQSLTRTVCAAGCARGKPLPTDCHEQC 240
Db      181 LTLIDNRSRACHPCSPMCKGSRGWGSESSDCQSLTRTVCAAGCARGKPLPTDCHEQC 240

```

Search completed: July 22, 2003, 08:41:51  
Job time : 44.1589 secs

```
QY 241 AAGCTGPKHSIDCLAHFNHSGICELHCPALVTYNTDTEESMPNEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSIDCLAHFNHSGICELHCPALVTYNTDTEESMPNEGRYTFGASCVTACP 300
QY 301 VNYLSTDVSGCTLVCPHNOEYTAEDGTORCEKSGPCARVCYGLKMOYIKANSKFIGIT 360
Db 301 VNYLSTDVSGCTLVCPHNOEYTAEDGTORCEKSGPCARVCYGLKMOYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVFETLEEITGYLISAMPDSL 420
Db 361 IGEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVFETLEEITGYLISAMPDSL 420
QY 421 DLSVFQNTQVIRGRIHNGAYSLTLQGLISWLGRLSLELGSGLAIHNTHTLCFVHTV 480
Db 421 DLSVFQNTQVIRGRIHNGAYSLTLQGLISWLGRLSLELGSGLAIHNTHTLSFVHTV 480
QY 481 PMDOLFBNPHQALHTANRPEDECEVCEGLACHQOLCARGHCWGPPTOCVNCQSFRLGQEC 540
Db 481 PMDOLFBNPHQALHTANRPEDECEVCEGLACHQOLCARGHCWGPPTOCVNCQSFRLGQEC 540
QY 541 VEECRVLOGLPREYVNAHQLPQHPECQPNQSVTCFGEADQVACAHYKDPPECVAC 600
Db 541 VEECRVLOGLPREYVNAHQLPQHPECQPNQSVTCFGEADQVACAHYKDPPECVAC 600
QY 601 PNNFTVSFWLVRPKVASHLE---PDEEGACOPCPINCTHSCVLDLDDKCGPAEQRAS 656
Db 601 PGG-----VKPDLSTYMPIMKFPDEEGACOPCPINCTHSCVLDLDDKCGPAEQRAS 650
QY 657 LTSIVSAVVGLLVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOA 716
Db 657 LTSIVSAVVGLLVVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNOA 710
QY 717 QMRILKETELRKVKVLGSGAFSTYKGIIPDGENVKIPVAIKVIRENTSPKANKEILDE 776
Db 711 QMRILKETELRKVKVLGSGAFSTYKGIIPDGENVKIPVAIKVIRENTSPKANKEILDE 770
QY 777 AYVAGVGSPIVSRLLIGICTSTVOLVTLQMPYGLDHYRENRGLSODLNMCMOIA 836
Db 771 AYVAGVGSPIVSRLLIGICTSTVOLVTLQMPYGLDHYRENRGLSODLNMCMOIA 830
QY 837 KMSYLEDVRLVRLDLAARNVLVKS PNHVKITDFGLARLDDIETEHADGKVPKXMA 896
Db 831 KMSYLEDVRLVRLDLAARNVLVKS PNHVKITDFGLARLDDIETEHADGKVPKXMA 890
QY 897 LESILRRRFTHQSDVMSYGVTVWELMTFGAKYDGIIPARETLDLEKGERLPQPTICTID 956
Db 891 LESILRRRFTHQSDVMSYGVTVWELMTFGAKYDGIIPARETLDLEKGERLPQPTICTID 950
QY 957 VYMIWKCMMIDSECRPRFRELVSSEFSRMARDPQRFVITONEDLGPAASPDLSTFYRSLLE 1016
Db 951 VYMIWKCMMIDSECRPRFRELVSSEFSRMARDPQRFVITONEDLGPAASPDLSTFYRSLLE 1010
QY 1017 DDDMGDDLVDAERYLVLPQGGFFCPDPAFGAGVTHHRSSSTRSGGDLTLGLEPSEBEA 1076
Db 1011 DDDMGDDLVDAERYLVLPQGGFFCPDPAFGAGVTHHRSSSTRSGGDLTLGLEPSEBEA 1070
QY 1077 PRSPFLAPSEBAGSDVFDGDLGKAAGLQSLPTHDPSPLOYSEDPVPLPSETDGYVAP 1136
Db 1071 PRSPFLAPSEBAGSDVFDGDLGKAAGLQSLPTHDPSPLOYSEDPVPLPSETDGYVAP 1130
QY 1137 LTCSPQPEYVNOQDVRFQPPSPFREGPLPAARPAAGATLERAKTLSPGXNGVYKDVAFGGA 1196
Db 1131 LTCSPQPEYVNOQDVRFQPPSPFREGPLPAARPAAGATLERAKTLSPGXNGVYKDVAFGGA 1190
QY 1197 VENPEYLTPOGGAAPQHPHPPAFSPAFNDLVYMDODPERGAPSTFKGTPTAENPEYLG 1256
Db 1191 VENPEYLTPOGGAAPQHPHPPAFSPAFNDLVYMDODPERGAPSTFKGTPTAENPEYLG 1250
QY 1257 LDVPV 1261
Db 1251 LDVPV 1255
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

# OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.2855 Seconds

(without alignments)  
5413.772 Million cell updates/sec

Title: SEQ4-632-652-14

Perfect score: 6834

Sequence: 1 MELAALCRMGALLALLPPCA.....TRKPTTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: p1r1:\*\*

2: p1r2:\*\*

3: p1r3:\*\*

4: p1r4:\*\*

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

Result No.	Score	Query Match	Length	DB ID	Description
1	6623	96.9	1255	1 A24571	protein-tyrosine k
2	5864	85.8	1260	1 TVRNU	protein-tyrosine k
3	5841.5	85.5	1254	1 I48161	p-185 precursor
4	3113	45.6	1210	1 GQHDE	epidermal growth f
5	3080	45.1	1210	1 A53183	epidermal growth f
6	3049.5	44.6	1223	1 TVCHLV	epidermal growth f
7	2920.5	42.7	1308	2 A47253	epidermal growth f
8	2627	38.4	1166	1 S06142	protein-tyrosine k
9	2389.5	35.0	1342	2 A36223	kinase-related tra
10	2302.5	33.7	1339	2 JCA487	epidermal growth f
11	1721.5	25.2	698	1 TVFVLV	epidermal growth f
12	1658	24.3	604	1 TVYOH	protein-tyrosine k
13	1611.5	23.6	1330	1 GQFPE	epidermal growth f
14	1602	23.4	544	2 S35745	protein-tyrosine k
15	1595	23.3	545	2 S00727	kinase-related tra
16	1578	23.1	540	2 B44726	protein-tyrosine k
17	1576	23.1	540	1 TVFVEB	epidermal growth f
18	1509	22.1	644	1 A36325	epidermal growth f
19	1274	18.6	1333	2 E88257	protein-tyrosine k
20	1274	18.6	1374	2 S70712	protein-tyrosine k
21	1195	17.5	1369	2 S70713	protein-tyrosine k
22	1150	16.8	1717	1 A45558	epidermal growth f
23	1136	16.5	527	2 A42032	epidermal growth f
24	975.5	14.3	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	713	10.4	1363	2 T43220	insulin-like growth
28	694	10.2	1382	1 INHUR	insulin receptor p
29	693.5	10.1	1372	2 A34157	insulin receptor p

30	690	10.1	1383	2 A36080	insulin receptor p
31	680	10.0	1477	2 T18534	protein-tyrosine k
32	680	10.0	1607	2 T43212	insulin-like growth
33	679.5	9.9	1300	2 A36502	insulin receptor-r
34	669	9.8	1268	2 B36502	insulin receptor-r
35	644	9.4	1367	1 IGHURL	insulin-like growth
36	629	9.2	1371	2 A33837	insulin-like growth
37	622.5	9.1	2148	1 A56081	insulin receptor -
38	617	9.0	1390	2 T30346	insulin receptor -
39	616	9.0	2101	2 S57245	insulin receptor (
40	591	8.6	1114	1 S05582	protein-tyrosine k
41	589.5	8.6	1028	2 A45388	protein-tyrosine k
42	586	8.6	1091	2 S33596	protein-tyrosine k
43	579	8.5	1052	2 A46166	protein-tyrosine k
44	578.5	8.5	952	2 I50612	protein-tyrosine k
45	576.5	8.4	984	2 A39753	protein-tyrosine k

## ALIGNMENTS

### RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N.Altimate names: c-erbB-2 protein precursor; kinase-related transforming protein

C/Species: Homo sapiens (man)

C/Date: 25-Oct-1987 #sequence\_revision 06-Dec-1996 #text\_change 11-Jun-1999

C/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human c-erbB-2 gene to epidermal gro

A/Reference number: A24571; MUID:86118663; PMID:3003577

A/Accession: A24571

A/Molecule type: mRNA

A/Residues: 1-1255 <YAM>

A/Cross-references: GB:X03363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T. Proc Natl. Acad Sci U.S.A. 82, 6497-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/ep

A/Reference number: A25491; MUID:86016729; PMID:2995967

A/Accession: A25491

A/Molecule type: DNA

A/Residues: 737-1031 <SEM>

A/Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R/Coussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seebu Science 230, 1132-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares ch

A/Reference number: A44188; MUID:86070181; PMID:2999974

A/Accession: A44188

A/Molecule type: DNA

A/Residues: 740-910 <COU>

A/Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A/Accession: B44188

A/Molecule type: mRNA

-12 549 3 993" data-label="Text">

A/Residues: 1-517, 'RALU', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU>

-35 549 4 993" data-label="Text">

A/Cross-references: GB:M11730; NID:G183986

-58 549 5 993" data-label="Text">

R/King, C.R.; Kraus, M.H.; Aaronson, S.A. Science 229, 974-976, 1985

-81 549 6 993" data-label="Text">

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

-104 549 7 993" data-label="Text">

A/Reference number: I59509; MUID:85272597; PMID:2992089

-127 549 8 993" data-label="Text">

A/Accession: I59509

-150 549 9 993" data-label="Text">

A/Status: translated from GB/EMBL/DBJ

-173 549 10 993" data-label="Text">

A/Molecule type: DNA

-196 549 11 993" data-label="Text">

A/Residues: 832-909 <REX>

-219 549 12 993" data-label="Text">

A/Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

-242 549 13 993" data-label="Text">

R/Itai, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D. Mol. Cell. Biol. 7, 2597-2601, 1987

-265 549 14 993" data-label="Text">

A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcript;

-288 549 15 993" data-label="Text">

A/Reference number: I57622; MUID:87286898; PMID:3039351

-311 549 16 993" data-label="Text">

A/Accession: I57622

-334 549 17 993" data-label="Text">

A/Status: translated from GB/EMBL/DBJ

-357 549 18 993" data-label="Text">

A/Molecule type: DNA

-380 549 19 993" data-label="Text">

A/Residues: 1-191 <TAU>

A:Cross-references: GB:M16792; NID:g183983; PIDN:AAA5637.1; PID:g553332  
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30  
 C:Genetics:  
 A:Gene: GDB:ERBB2; NGI; NEU; HER-2  
 A:Cross-references: GDB:120613; OMIM:164870  
 A:Map position: 17q21.1-17q21.1  
 A:Introns: 25/1; 75/3; 147/1; 883/3  
 A>Note: The list of introns is incomplete  
 C:Function:  
 A:Description: catalyzes the phosphorylation of a peptide tyrosine residue by ATP  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:22-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>  
 F:22-653/Domain: extracellular #status predicted <EXT>  
 F:70-304/Domain: EGF receptor extracellular domain repeat <EB1>  
 F:395-605/Domain: EGF receptor extracellular domain repeat <EB2>  
 F:654-675/Domain: transmembrane #status predicted <TM>  
 F:676-1255/Domain: intracellular #status predicted <INT>  
 F:718-993/Region: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif  
 F:68,124,187,259,530,571,629/Binding site: carboxylate (asn) (covalent) #status predicted  
 F:686/Binding site: phosphate (thr) (covalent) (by protein kinase C) #status predicted  
 F:753/Active site: Lys #status predicted  
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 96.9% Score 6623; DB 1; Length 1255;  
 Best local similarity 97.3% Pred. No. 3.6e-25;  
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLLALPPGAASGVCTGTDMKRLPASPETHLDMLSHLYOGCCVVOGNT 60  
 DB 1 MELAALCRWGLLLALPPGAASGVCTGTDMKRLPASPETHLDMLSHLYOGCCVVOGNT 60  
 QY 61 ELTLPNTASLSFODIOENOGVYLIAHQVRLQRLRYRGTCQFEDNYALAVDNG 120  
 DB 61 ELTLPNTASLSFODIOENOGVYLIAHQVRLQRLRYRGTCQFEDNYALAVDNG 120  
 QY 121 DPLNNTFVGTASFGGLRELQRLTEILKGVLIQRPOLCYODTILMKDIFKNNQLA 180  
 DB 121 DPLNNTFVGTASFGGLRELQRLTEILKGVLIQRPOLCYODTILMKDIFKNNQLA 180  
 QY 181 LTLIDTNRSPACHPCSPCKSKRCWSESSDDCSITRTVCAGGACRKGPLPTDCHEQC 240  
 DB 181 LTLIDTNRSPACHPCSPCKSKRCWSESSDDCSITRTVCAGGACRKGPLPTDCHEQC 240  
 QY 241 AAGCTGPRGSDCLACLNHNSGICELHCPALVTNTDFEEMPRPEGRYTGASCTYACR 300  
 DB 241 AAGCTGPRGSDCLACLNHNSGICELHCPALVTNTDFEEMPRPEGRYTGASCTYACR 300  
 QY 301 YNYLSTDVSGCTLVCPHNGEVTADGTQRCCKSKPCARVCYGLNQYIKANSKFTGIT 360  
 DB 301 YNYLSTDVSGCTLVCPHNGEVTADGTQRCCKSKPCARVCYGLNQYIKANSKFTGIT 360  
 QY 361 YNYLSTDVSGCTLVCPHNGEVTADGTQRCCKSKPCARVCYGLNQYIKANSKFTGIT 360  
 DB 361 YNYLSTDVSGCTLVCPHNGEVTADGTQRCCKSKPCARVCYGLNQYIKANSKFTGIT 360  
 QY 421 DLSVFNQVIRGRILHNGAYSLTQGLIGISWGLRSLRELGSGLALIHNTHLCPHYTV 480  
 DB 421 DLSVFNQVIRGRILHNGAYSLTQGLIGISWGLRSLRELGSGLALIHNTHLCPHYTV 480  
 QY 481 PMDQLFRPHQALLHTANRPEDECVGEGELACHQICARHCHGPRPTQVNCOSQLRQEC 540  
 DB 481 PMDQLFRPHQALLHTANRPEDECVGEGELACHQICARHCHGPRPTQVNCOSQLRQEC 540  
 QY 541 VEEGRVGLGLPREVYNAHCLPCHPECCQPNQSVTCEPEADQCVACAHYDPPFCVARC 600  
 DB 541 VEEGRVGLGLPREVYNAHCLPCHPECCQPNQSVTCEPEADQCVACAHYDPPFCVARC 600  
 QY 601 PSQVPRDLSTYPMKPEDEGACQPCPNCNCHSCVDLDDKCCPAPORASPLTISFNFTV 660  
 DB 601 PSQVPRDLSTYPMKPEDEGACQPCPNCNCHSCVDLDDKCCPAPORASPLTISFNFTV 660

DB 601 PSQVPRDLSTYPMKPEDEGACQPCPNCNCHSCVDLDDKCCPAPORASPLTISAVVG 660  
 QY 661 SFMLRVPKVSASHLEKRRQCKIRKTYMRLLLOETELVEPLTSSGAMPNDAQRIKLTETL 720  
 DB 661 SFMLRVPKVSASHLEKRRQCKIRKTYMRLLLOETELVEPLTSSGAMPNDAQRIKLTETL 720  
 QY 721 RKVKVLSGAFGVVKGVIPIPGENVKIPVAIKVIRENTPANKKEILDEAVYAGVSGP 780  
 DB 721 RKVKVLSGAFGVVKGVIPIPGENVKIPVAIKVIRENTPANKKEILDEAVYAGVSGP 780  
 QY 781 YVSRLLGICLTSTVQLVTLQMPYGCCLDVRNRRGLSGQDILNMCQIAKMSYLEVR 840  
 DB 781 YVSRLLGICLTSTVQLVTLQMPYGCCLDVRNRRGLSGQDILNMCQIAKMSYLEVR 840  
 QY 841 LVHRDLAARNVLYKSNHVKINDFGIARLLIDETEVADGGKVPKMMALSLIRRRFT 900  
 DB 841 LVHRDLAARNVLYKSNHVKINDFGIARLLIDETEVADGGKVPKMMALSLIRRRFT 900  
 QY 901 HOSDWSYGVTVWELMTFGAKPYDGIIPAREIDLEKGERLPQPPICITIDVYIMVKCM 960  
 DB 901 HOSDWSYGVTVWELMTFGAKPYDGIIPAREIDLEKGERLPQPPICITIDVYIMVKCM 960  
 QY 961 IDSECRPRRELVSERFARMARDPQRFVYQNEIDGPAFLDSTFRRSLLEDMDGLVVA 1020  
 DB 961 IDSECRPRRELVSERFARMARDPQRFVYQNEIDGPAFLDSTFRRSLLEDMDGLVVA 1020  
 QY 1021 EEVLVQGGFFCPDPAPGAGWVHRHRSSTRSGGDDTLGLPESEEPAPSPPLAPSEG 1080  
 DB 1021 EEVLVQGGFFCPDPAPGAGWVHRHRSSTRSGGDDTLGLPESEEPAPSPPLAPSEG 1080  
 QY 1081 AGSDVFDGLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYVAPLTCSPQEVY 1140  
 DB 1081 AGSDVFDGLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYVAPLTCSPQEVY 1140  
 QY 1141 NQPDVPRPSPREGLPAPRPAAGATLEPAKTLSPCKNVWDVFAFGAVNPEVLTQ 1200  
 DB 1141 NQPDVPRPSPREGLPAPRPAAGATLEPAKTLSPCKNVWDVFAFGAVNPEVLTQ 1200  
 QY 1201 GGAAPQHPHPAPSPAFDNLVYWDQPPERGAPEPTFKTPTAENPEVGLDVPV 1255  
 DB 1201 GGAAPQHPHPAPSPAFDNLVYWDQPPERGAPEPTFKTPTAENPEVGLDVPV 1255

RESULT 2  
 TVRNTU  
 protein-tyrosine kinase (EC 2.7.1.112) neu precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 11-Jun-1999  
 C:Accession: A24562; A61204  
 R:Bergmann, C.I.; Hung, M.C.; Weinberg, R.A.  
 Nature 319, 226-230, 1986  
 A:Title: The neu oncogene encodes an epidermal growth factor receptor-related prote  
 A:Reference number: A24562; MUID:86118662; PMID:3945311  
 A:Accession: A24562  
 A:Molecule type: mRNA  
 A:Residues: 11260 <BAR>  
 A:Cross-references: EMBL:X03362; NID:g56745; PIDN:CAA27059.1; PID:g56746  
 R:Masu, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cc  
 Carcinogenesis 12, 1975-1978, 1991  
 A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals  
 2-thiathiolyl]formamide or N-methyl-N-nitrosourea.  
 A:Reference number: A61204; MUID:92035293; PMID:1682063  
 A:Accession: A61204  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 637-663; 'V', 665-702 <MAS>  
 A>Note: authors translated the codon GCA for residue 25 as Val  
 C:Genetics:  
 A:Gene: neu  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; ph  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

F:658-660/Domain: transmembrane #status predicted <TM>  
 F:723-968/Domain: protein kinase homology <KIN>  
 F:731-739/Region: protein kinase ATP-binding motif  
 F:711,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:691/Binding site: phosphate (Thr) (covalent) #status predicted  
 F:758/Active site: lys #status predicted  
 F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.8%; Score 5864; DB 1; Length 1260;  
 Best Local Similarity 86.2%; Pred. No. 5,7e-234;  
 Matches 1084; Conservative 48; Mismatches 123; Indels 2; Gaps 2;

QY 1 MELALCRWGLLALLPPGASTVCTGTDMLRLPASPEHLDMRLHLYOGCCVYVQNL 60  
 DB 4 MELAMCRWGLLALLPPGAGTGTCTGTDMLRLPASPEHLDMRLHLYOGCCVYVQNL 63  
 QY 61 ELTYLPTNASISFLDDIOEVQGYVLIANOVQVPLRLRIVRGQLFEDNYALAVLDNG 120  
 DB 64 ELTYVPANASISFLDDIOEVQGYVLIANOVQVPLRLRIVRGQLFEDNYALAVLDNR 123  
 QY 121 DPLNNTPTVT-GASPGRLRLQLRSLTEILKGGVLIQRNPOLCYODTILMKDI PHKNNOL 179  
 DB 124 DPODNVAASTPGRTRPEGLRLQLRSLTEILKGGVLIQRNPOLCYODVILKDYFRKNNOL 183  
 QY 180 ALTLIDTNRSPACHPSPMCKGSRCWGESSEDCOSLTRVACAGSCARCKGLPTDCHEQ 239  
 DB 184 APVIDITNRSPACHPPACKCKDNHGWGESSEDCOILITGTCGACARCKGLPTDCHEQ 243  
 QY 240 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTAC 299  
 DB 244 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTTC 303  
 QY 300 PYNLTSDVGSCTIVCPLHNOEVTABDGTORCEKSKPCARVCYGLGMQYIKANSKFIGI 359  
 DB 304 PYNLTSTEVGSCTIVCPNNOEVTABDGTORCEKSKPCARVCYGLGMELHGRARAITSQ 363  
 QY 360 TELEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVETLEETITGLYISAMPDGL 419  
 DB 364 NVOEFQCKKIFGSLAFLPESFDGDPSSGIALPPELOVETLEETITGLYISAMPDGL 423  
 QY 420 PDLSVFONLQVIRGRILHNGAVSLTLOGISLWGLRSLRELGSGALIHNTHLCFVHT 479  
 DB 424 RDLVSFONLRIIRGRILHNGAVSLTLOGISLHSLRSLRELGSGALIHNTHLCFVHT 483  
 QY 480 VPMDOLFERNPHQALLHNPANRPEDE-CYGEGLACHQICARHCHGCPPTQCVNCSOTLRQ 538  
 DB 484 VPMDOLFERNPHQALLHNSGNRPEEDLCVSSGLVCNSLCAHGHGCPPTQCVNCSOTLRQ 543  
 QY 539 ECVEECRVLQGLFREYVNAHCLPCHPECCQPNQSVYTCFEPADQCACAHYKDPFCVA 598  
 DB 544 ECVEECRVMKGLPREYVSDKRLPCHPECCQPNQSVSEFCSEADQCAAHYKDDSSCYA 603  
 QY 599 RCSSGVKPDLSYMPIMKFPDEBAGCQPCPINCNSCYDLDKCCPAPORASPLTIS-FNNF 658  
 DB 604 RCSSGVKPDLSYMPIMKFPDEBAGCQPCPINCNSCYDLDKCCPAPORASPLTIFATV 663  
 QY 659 TVSFMDLVPKVASHLERKROQKI RKYTMRLLOETELVEPLTPSGAMPNOAMRLKET 718  
 DB 664 EGYLLFLIVVVVGLIKRRQKIRKYTMRLLOETELVEPLTPSGAMPNOAMRLKET 723  
 QY 719 ELKRVKVLGSGAGTYKGIWIPDGENVKI PVAIKYLRNTSPKANKIILDEAYVAVG 778  
 DB 724 ELKRVKVLGSGAGTYKGIWIPDGENVKI PVAIKYLRNTSPKANKIILDEAYVAVG 783  
 QY 779 SPVYSRLGICLSTVQVLTQMLPYGLLDHVENNGRIGSODLWMQIAKMSYLED 838  
 DB 784 SPVYSRLGICLSTVQVLTQMLPYGLLDHVENNGRIGSODLWMQIAKMSYLED 843  
 QY 839 VRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPIMMALLESILRR 898  
 DB 844 VRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPIMMALLESILRR 903  
 QY 899 FTHQSDVWSYGVTVWELMTFGAKFPYDGIAPAREIPDLLEKGERLPPTICTIDVYIMLVKC 958

DB 904 FTHQSDVWSYGVTVWELMTFGAKFPYDGIAPAREIPDLLEKGERLPPTICTIDVYIMLVKC 963  
 QY 959 WMIDSCRRPFEIUSFSEFMRMDQRFVYVIONELGSPASPLDSTFYRSLLEDDDDGDLV 1018  
 DB 964 WMIDSCRRPFEIUSFSEFMRMDQRFVYVIONELGSPASPLDSTFYRSLLEDDDDGDLV 1023  
 QY 1019 DAEELYVPQGFECFCDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEEFAPRSLAS 1078  
 DB 1024 DAEELYVPQGFECFCDPAAGAGMHHRRSSSTRSGGDLTLGLEPSEEFAPRSLAS 1083  
 QY 1079 EGAGSDVPFGDILCMKAAKLGSLPHDPSFLORYSEDPVLPJSETDGYVAPLTCSPQE 1138  
 DB 1084 EGAGSDVPFGDILCMKAAKLGSLPHDPSFLORYSEDPVLPJSETDGYVAPLTCSPQE 1143  
 QY 1139 YYNOPRFPQPSPEBGLPAPRPAATLEAKTSLPGKNGVYKDVAFAGAVENPEYLT 1198  
 DB 1144 YYNQSEVQOPPLTPEBGLPAPRPAATLEAKTSLPGKNGVYKDVAFAGAVENPEYLT 1203  
 QY 1199 PQGAAPDPHPPEAPSPADNLYWDODPPERGAPSTFKGTPTAENPEYLGLDVPPV 1255  
 DB 1204 PREGTASPHPPAPSPADNLYWDQNSSEGGPPSPNEGTPTAENPEYLGLDVPPV 1260

## RESULT 3

148161  
 p-185 precursor - golden hamster  
 C:Species: Mesocricetus auratus (golden hamster)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Jun-1999  
 C:Accession: I48161  
 R:Nakamura, T.; Ushijima, T.; Iehizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishi  
 Gene 140, 251-255, 1994  
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.  
 A:Reference number: 148161; PMID:94193007; PMID:7908275  
 A:Accession: I48161  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1254 <RES>  
 A:Cross-references: GB:D16295; NID:9493236; PID:BA003801.1; PID:9747595  
 C:Genetics:  
 A:Gene: neu  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP  
 F:126-983/Domain: protein kinase homology <KIN>  
 F:126-734/Region: protein kinase ATP-binding motif

Query Match 85.5%; Score 5841.5; DB 2; Length 1254;  
 Best Local Similarity 85.7%; Pred. No. 4,7e-233;  
 Matches 1076; Conservative 57; Mismatches 121; Indels 1; Gaps 1;

QY 1 MELALCRWGLLALLPPGASTVCTGTDMLRLPASPEHLDMRLHLYOGCCVYVQNL 60  
 DB 1 MELAMCRWGLLALLPPGASTVCTGTDMLRLPASPEHLDMRLHLYOGCCVYVQNL 60  
 QY 61 ELTYLPTNASISFLDDIOEVQGYVLIANOVQVPLRLRIVRGQLFEDNYALAVLDNG 120  
 DB 61 ELTYVPANASISFLDDIOEVQGYVLIANOVQVPLRLRIVRGQLFEDNYALAVLDNR 120  
 QY 121 DPLNNTPTVT-GASPGRLRLQLRSLTEILKGGVLIQRNPOLCYODTILMKDI PHKNNOL 180  
 DB 121 DPLNNTPTVT-GASPGRLRLQLRSLTEILKGGVLIQRNPOLCYODTILMKDI PHKNNOL 180  
 QY 181 LTLIDTNRSPACHPSPMCKGSRCWGESSEDCOSLTRVACAGSCARCKGLPTDCHEQ 240  
 DB 181 LTLIDTNRSPACHPSPMCKGSRCWGESSEDCOSLTRVACAGSCARCKGLPTDCHEQ 240  
 QY 241 PVDIDTNRSPACHPSPMCKGSRCWGESSEDCOSLTRVACAGSCARCKGLPTDCHEQ 240  
 DB 241 PVDIDTNRSPACHPSPMCKGSRCWGESSEDCOSLTRVACAGSCARCKGLPTDCHEQ 240  
 QY 241 AAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTAC 300  
 DB 241 AAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTAC 300  
 QY 301 YNYLSTDVGSCTIVCPLHNOEVTABDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360  
 DB 301 YNYLSTDVGSCTIVCPLHNOEVTABDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360  
 QY 360 YNYLSTDVGSCTIVCPLHNOEVTABDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360  
 DB 360 YNYLSTDVGSCTIVCPLHNOEVTABDGTORCEKSKPCARVCYGLGMQYIKANSKFIGIT 360

```

QY 361 ELERAGCKXIFGSLAFIPESFDGDPASNTAPLOEOLQVETTEITGYLYISAMPDLSL 420
DB 361 IQERAGCKXIFGSLAFIPESFDGDPASNTAPLOEOLQVETTEITGYLYISAMPDLSL 420
QY 421 DLSVFNOLVIRGRILHNAGAYSLTLOGISWLGRLSRLREISGLALIHNTHLCEVHTV 480
DB 421 DLSVFNOLVIRGRILHNAGAYSLTLOGISWLGRLSRLREISGLALIHNTHLCEVHTV 480
QY 481 PMDQFRPHQALLHTANRPEDECVSEGLACHQICAGHCRCRPPPTQCVNCSQTLRQEC 540
DB 481 PMDQFRPHQALLHTANRPEDECVSEGLACHQICAGHCRCRPPPTQCVNCSQTLRQEC 540
QY 541 VEECRVLOGLBREYVNAHCLPCHPECOPQNGSTCEPPEADQCVAAHYXDPFCVARC 600
DB 541 VKECRVMKGLPREVYNGHCLPCHPECOPQNGSTCEPPEADQCVAAHYXDPFCVARC 600
QY 601 PSQVYKPDLSYMPIMKYPDEEGMCPQPCINCHSCVDDDERGCPAECQASPATSIATVVG 660
DB 601 PSQVYKPDLSYMPIMKYPDEEGMCPQPCINCHSCVDDDERGCPAECQASPATSIATVVG 660
QY 661 SFMLRVPKVSASHLEKROCKIRKYMRLLOETELVEPLTPSGAMFNQOMRLKETEL 720
DB 661 ILFLVIGVAVVGLIKRRQKIRKYMRLLOETELVEPLTPSGAMFNQOMRLKETEL 720
QY 721 RKVYVLSGAFGYVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVYVAGVSP 780
DB 721 RKVYVLSGAFGYVYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVYVAGVSP 780
QY 781 YVSRLLGLCTSTVOLVQLMPYGLLDHVENENGRISQDOLMWCQIAKMSYLEDVR 840
DB 781 YVSRLLGLCTSTVOLVQLMPYGLLDHVENENGRISQDOLMWCQIAKMSYLEDVR 840
QY 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPIMKMALESILRRFT 900
DB 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPIMKMALESILRRFT 900
QY 901 HQSDVMSGYTVVLMFTFGAKFYDGIPTREIPDLLENGERLPOPPICITIDVYMMVKCM 960
DB 901 HQSDVMSGYTVVLMFTFGAKFYDGIPTREIPDLLENGERLPOPPICITIDVYMMVKCM 960
QY 961 IDSECRPFRELVSEFSMARDPQRFVVIQNEIDGPASPLDSTFYRSLLEDDEMDGLVDA 1020
DB 961 IDSECRPFRELVSEFSMARDPQRFVVIQNEIDGPASPLDSTFYRSLLEDDEMDGLVDA 1020
QY 1021 EBYVLPQGGFCPPDPAPAGAMVHRHRSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080
DB 1021 EBYVLPQGGFCPPDPAPAGAMVHRHRSSTRSGGDLTLGLEPSEEBAPRSLAPSEG 1080
QY 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLQRYSEDPVPLPSETDGVVAPLTCGPOREYV 1140
DB 1081 AGSDVFDGDLGMAKAGLQSLPTHDPSPLQRYSEDPVPLPSETDGVVAPLTCGPOREYV 1140
QY 1141 NQPDVRRPQPPSFRGGLPAPAPAGATLERAKTISPGKNGVYKDVAFAGAVENEYLTPO 1200
DB 1141 NQPDVRRPQPPSFRGGLPAPAPAGATLERAKTISPGKNGVYKDVAFAGAVENEYLTPO 1200
QY 1201 GGAAPQPPHPPAPFAPDNTLYYMDODPBERGAPSTFKGTAAENPYLG.DVYV 1255
DB 1201 GGAAPQPPHPPAPFAPDNTLYYMDODPBERGAPSTFKGTAAENPYLG.DVYV 1255
QY 1255 GGSASQPH-PALCPAPDNLVYVMDQDSERGSPPNTEGTPTAENPYLG.DVYV 1254
DB 1255 GGSASQPH-PALCPAPDNLVYVMDQDSERGSPPNTEGTPTAENPYLG.DVYV 1254

```

## RESULT 4

GQOHUE  
 epidermal growth factor receptor precursor - human  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbb  
 C:Species: Homo sapiens (man)  
 C:Date: 15-Nov-1994 #sequence\_revision 27-Nov-1995 #text\_change 11-Jun-1999  
 C:Accession: A00641, A25772, A38672, A00642, A43615, A23062, A05281, A60143, A33  
 R:Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; Y  
 rg, P.H.  
 A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of

A:Reference number: A00641; MUID:84219729; PMID:6328312  
 A:Accession: A00641  
 A:Molecule type: mRNA  
 A:Residues: 1-1210 <UL>  
 A:Cross-references: EMBL:X00588; NID:g31113; PIDN:CAA2540.1; PID:g757924  
 A:Note: the authors translated the codon AAG for residue 540 as Aen  
 R:Ulrich, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985  
 A:Title: Characterization and sequence of the promoter region of the human epiderma  
 A:Reference number: A25772; MUID:85270438; PMID:2991899  
 A:Accession: A25772  
 A:Molecule type: DNA  
 A:Status: translation not shown  
 A:Residues: 1-29 <IGH>  
 A:Cross-references: GB:M11234; NID:g181981; PIDN:AAA52370.1; PID:g553272  
 R:Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ulrich, A.; Waterfield, M.  
 Oncogene Res. 1, 375-396, 1987  
 A:Title: The human EGF receptor gene: structure of the 110 kb locus and identificat  
 A:Reference number: S30024; MUID:88217333; PMID:3329716  
 A:Accession: S30024  
 A:Molecule type: DNA  
 A:Residues: 1-29 <HA2>  
 A:Cross-references: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119  
 R:Haley, J.D.; Waterfield, M.D.  
 J. Biol. Chem. 266, 1746-1753, 1991  
 A:Title: Contributory effects of de Novo transcription and premature transcript tex  
 A:Reference number: A38672; MUID:91107677; PMID:1988448  
 A:Accession: A38672  
 A:Molecule type: DNA  
 A:Residues: 1-29 <HNL>  
 A:Cross-references: GB:M38425; NID:g181977; PIDN:AAA63171.1; PID:g553271  
 A:Experimental source: carcinoma cell line A431-7  
 R:Xu, Y.; Ishii, S.; Clark, A.J.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.  
 Nature 309, 806-810, 1984  
 A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of  
 A:Reference number: A00642; MUID:84245835; PMID:6330563  
 A:Accession: A00642  
 A:Molecule type: mRNA  
 A:Residues: 150-187, 'KSVIQAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 30  
 A:Residues: 'RCAMRRA', 150-187, 'KSVIQAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 30  
 A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF  
 R:Ulrich, C.R.; Chen, W.S.; Krutger, W.; Stolarsky, L.S.; Weber, M.; Evans, R.M.; Ver  
 Science 224, 843-848, 1984  
 A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplifica  
 A:Reference number: A43615; MUID:84196372; PMID:6326261  
 A:Accession: A43615  
 A:Molecule type: mRNA  
 A:Residues: 713-964 <LIN>  
 A:Experimental source: epidermoid carcinoma cell line A431  
 R:Stamen, F.A.; Gope, M.L.; Schulz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.  
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984  
 A:Reference number: A23062; MUID:85046483; PMID:6093780  
 A:Accession: A23062  
 A:Molecule type: mRNA  
 A:Residues: 1028-1210 <SIM>  
 R:Weber, M.; Gill, G.N.; Speiser, J.  
 Science 224, 294-297, 1984  
 A:Reference number: A05281; MUID:84172183; PMID:6324343  
 A:Accession: A05281  
 A:Molecule type: protein  
 A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>  
 R:Russo, M.W.; Lukac, T.J.; Cohen, S.; Staros, J.V.  
 J. Biol. Chem. 260, 5205-5208, 1985  
 A:Title: Identification of residues in the nucleotide binding site of the epidermal  
 A:Reference number: A60143; MUID:85182850; PMID:2985860  
 A:Accession: A60143  
 A:Molecule type: protein  
 A:Residues: 740-744, 'X', 746-747 <RUS>  
 R:Woczkowski, B.; Mosig, G.; Cohen, S.  
 Nature 309, 270-273, 1984  
 A:Title: ATP-stimulated interaction between epidermal growth factor receptor and su  
 A:Reference number: A38023; MUID:84191554; PMID:6325948  
 A:Contents: annotation; receptor activity



A>Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA  
 R;Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsch, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.  
 Cell 59, 33-43, 1989  
 A>Title: Functional independence of the epidermal growth factor receptor from a domain I  
 A:Reference number: A33331; MIMD:900032233; PMID:21790960  
 A:Content: annotation; internalization signal  
 C/Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor  
 C/Genetics:  
 A:Gene: GDB:EGFR  
 A:Cross-references: GDB:120610; OMIM:131550  
 A:Map position: 7p12.3-7p12.1  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 F:124/Domain: signal sequence #status predicted <SIG>  
 F:25-1210/Product: EGF receptor #status predicted <MAT>  
 F:25-645/Domain: extracellular #status predicted <EXT>  
 F:75-300/Domain: EGF receptor extracellular domain repeat <E1>  
 F:390-600/Domain: EGF receptor extracellular domain repeat <E2>  
 F:669-1210/Domain: transmembrane #status predicted <TM>  
 F:710-975/Domain: intracellular #status predicted <INT>  
 F:718-926/Region: protein kinase homology <KIN>  
 F:999-1046/Region: coated-pit mediated internalization signal  
 F:1047-1210/Region: inhibitory  
 F:128,175,352,413,444,528,603/Binding site: carboxydrate (Asn) (covalent) #status predic  
 F:745/Active site: Lys #status experimental

Query Match 45.6%; Score 3113; DB 1; Length 1210;  
 Best Local Similarity 49.5%; Pred. No. 6,3e-121;  
 Matches 625; Conservative 173; Mismatches 363; Indels 102; Gaps 21;

```

QY 11 LLLALLPFGNA--STQVCGTDMKLLPASPETHLDMRLKLYGCGVQGNLETLPLTN 68
DB 14 LLALPASPALFEKKYVCGSTNKLTLQGFEDHFLSLQMFNCCVVGNETTYQEN 73
QY 69 ASLSFLQDIOEVGVYLAHQVROVPLRLIRIVRGTOLEFEDVYALVDNGDPLNTRP 128
DB 74 YDSFLKTIQEVAGVYLAINTVERIPLENLQILRNMYENSYALAVLSND----- 126
QY 129 VTASAPGRLRLSLTELKGGVLIQRNPOLCYDITIMKDIFFKNNQALATLIDTR 188
DB 127 ---ANKTGLKELMRNLOELHGAVFNSNPALCNVESIQMDIVSDLSNMSDFQNH 183
QY 189 SRACHPCSPWCKSGRCSESEDCOSLTRVCGAGA-RCKGFLPTDCCHEOCACCTGP 247
DB 184 LGSCKQKDBSCPRGSCWAGENCQKTLKICQCSGRCKGSPDCCCHNQCAGCTGP 243
QY 248 KHSDDLACLFHNSGICEIHCPLVYNTDTFESMNPBGRYTFGASCVYACPYNYLSTD 307
DB 244 RESDCLVCRKFRDEATCKDTCPLMLYNPPTYQMDVNPBGKYSFGATCVKCPBNVVD 303
QY 308 VGSCTLVCPLNHDEVTAEDGTORCEGCKSPCARVCGVGLQMOYIKANSKFTIGTELE-FAG 366
DB 304 HSGCVRAAGADSTEM-EDGCVKCKCKGECGRVCGIGGEFR-DLSINANTIKHFN 361
QY 367 CKKIFGSLFLPSPDGPASNTAPLQPEOLQVFELEETITGLYLSAMPDLSPLDVQ 426
DB 362 CTSISGDLHLPAVAFGDSFTTHPPDPEGLDILKTVKEITGLILQAMPENRTDHAEE 421
QY 427 NLOVITGRILHNAYSLTLQGLISMLGLRLSELGSLALIHNTLHCFVHTVPDOLF 486
DB 422 NEETIGRTKQHQQLAVAVSLNTSLGRSLKEISDGVIIISGNKLQYANTIMKKLF 481
QY 487 RNHQALHTANREDECVGEGLACHQLCARGHCWPGPTQCNQCSQFLRGECEYECV 546
DB 482 GSGGQKTKIISNNGENSKATGVCALCSPEGCWPERDPCVSCNVRGSECDKCKL 541
QY 547 LQGLPREYVNAHCLPCHPECPQNGSVTCFGEADQCVACAHYKDPFCVAPSGVXP 606
DB 542 LEEGPREFVENSECICHCECLQFQANNITCTGKGPNCITQCHAYIDGPHCVTCFAGVWG 601
QY 607 DLSYPMKFPDEGACQCPFINCTHSQVDLDKGPAPGRASPLTISINNTVSWFLRV 666

```

DB 602 ENNTL-VKTYADAGVCHLCHPNCYGGTGBGLESCPTNGKIP--SIATGVGALLLL 658  
 QY 667 PVASASHLERQCKIRKYTYMRRLQELVEIPRLPSGAMPQAMRLIKETELKRVYL 726  
 DB 659 VVALGILPMRRRIYRKETLRLLQELVEIPRLPSGAPQALRLIKETEFKIKYL 718  
 QY 727 GSGAGTYVKGWIPDGENVKIPVAIKYLRNTSPKANKELDEAYVAVGSPVYSRL 786  
 DB 719 GSGAGTYVKGWIPDGENVKIPVAIKYLRNTSPKANKELDEAYVAVGSPVYSRL 778  
 QY 787 GICLSTVQLTYQLMPYGLDLNHYENGRIGSODLLWCMQIANGSYLDEYLVHDL 846  
 DB 779 GICLSTVQLTYQLMPYGLDLNHYENGRIGSODLLWCMQIANGSYLDEYLVHDL 838  
 QY 847 AARNLVSPNHKKTIDFGARLDIDETRYHADGKVPKMMALLESILRRFTQSDVW 906  
 DB 839 AARNLVTPQWKTIDFGARLDIDETRYHADGKVPKMMALLESILRRFTQSDVW 898  
 QY 907 SYGVTVWELMTFGAKPYDGIAPAEIPDLLEKGERLPQPICTIDYVMIMVCMIDSECR 966  
 DB 899 SYGVTVWELMTFGSKPYDGIAPAEISISLEKGERLPQPICTIDYVMIMVCMIDADS 958  
 QY 967 PRERELVEEPMARDPQRFVIO-NEDLGASPLDSFYSLSLEDDCMGLVDAEYLV 1025  
 DB 959 PRERELVEEPMARDPQRFVIO-NEDLGASPLDSFYSLSLEDDCMGLVDAEYLV 1018  
 QY 1026 PQGFPCDPAFGAGMHHRRSSSTRSGGADLLTGLPESEEARSPPLAPSEAGSDV 1085  
 DB 1019 PQGFPCDPAFGAGMHHRRSSSTRSGGADLLTGLPESEEARSPPLAPSEAGSDV 1043  
 QY 1086 PDGDLGMAGKGLSLPETHDSPLQRYSEDFTVLPST--DGYVAPLTCSGPQRYVNP 1143  
 DB 1044 NSTVACIDRNGLOSPFKEDSFLQRYSSDTGALTEDSIDDTFL-----PVPEYINQ- 1095  
 QY 1144 DVREPSPREPGPPLAARPACTLEAKTLSPKQGVYKDVFAFGAVENPEVY-TPQG 1202  
 DB 1096 SVKRPASVQNPVHNPVLP-----APSRPHQD--PHSTAVANGELYNTVQ-- 1143  
 QY 1203 AARPPPPAPSPAFDNLVYDQ-----DP-----PERGAPSTFKGPTAENP 1246  
 DB 1144 -----PVCVNSTPDSBAHWAQKSHQISLDNPDYQDFFPKAKPRGIFKGS-TANNA 1195  
 QY 1247 EYL 1249  
 DB 1196 EYL 1198

RESULT 5  
 A53183  
 A:epidermal growth factor receptor precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999  
 C:Accession: A53183; A43818; S24942; A28941; S45325; 19643  
 R:Luetteke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N  
 Genes Dev. 9, 399-413, 1994  
 A>Title: The mouse waved-2 phenotype results from a point mutation in the EGF recep  
 A:Reference number: A53183; MIMD:94170986; PMID:8125255  
 A:Accession: A53183  
 A:Molecule type: mRNA  
 A:Residues: 1-1210 <LUE>  
 A:Cross-references: GB:U03425  
 R:Aviyl, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.  
 Oncogene 6, 673-676, 1991  
 A>Title: Comparison of EGF receptor sequences as a guide to study the ligand binding  
 A:Reference number: A43818; MIMD:91232866; PMID:2030916  
 A:Accession: A43818  
 A:Molecule type: mRNA  
 A:Residues: 1-714 <AVI>  
 A:Cross-references: GB:X59698  
 R:Steininger, D.P.; Serrero, G.  
 Submitted to the EMBL Data Library, June 1992  
 A:Reference number: S24942  
 A:Accession: S24942

A: Molecule type: mRNA  
 A: Residues: 969-971, 'K', 973-1115, 'D' <EIS>  
 A: Cross-references: EMBL:212608  
 R: Heilmann, G.U.; Gill, G.N.  
 J. Biol. Chem. 263, 13152-13159, 1988  
 A: Title: Epidermal growth factor receptor threonine and serine residues phosphorylated  
 A: Reference number: A28941; MUID: 88330814; PMID: 3138233  
 A: Accession: A28941  
 A: Molecule type: Protein  
 A: Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009, 'R', 1100-1101, 'D', 1102-1103, 'A', 1104-1105, 'S', 1106-1107, 'S', 1108-1109, 'S', 1110-1111, 'S', 1112-1113, 'S', 1114-1115, 'S', 1116-1117, 'S', 1118-1119, 'S', 1120-1121, 'S', 1122-1123, 'S', 1124-1125, 'S', 1126-1127, 'S', 1128-1129, 'S', 1130-1131, 'S', 1132-1133, 'S', 1134-1135, 'S', 1136-1137, 'S', 1138-1139, 'S', 1140-1141, 'S', 1142-1143, 'S', 1144-1145, 'S', 1146-1147, 'S', 1148-1149, 'S', 1150-1151, 'S', 1152-1153, 'S', 1154-1155, 'S', 1156-1157, 'S', 1158-1159, 'S', 1160-1161, 'S', 1162-1163, 'S', 1164-1165, 'S', 1166-1167, 'S', 1168-1169, 'S', 1170-1171, 'S', 1172-1173, 'S', 1174-1175, 'S', 1176-1177, 'S', 1178-1179, 'S', 1180-1181, 'S', 1182-1183, 'S', 1184-1185, 'S', 1186-1187, 'S', 1188-1189, 'S', 1190-1191, 'S', 1192-1193, 'S', 1194-1195, 'S', 1196-1197, 'S', 1198-1199, 'S', 1200-1201, 'S', 1202-1203, 'S', 1204-1205, 'S', 1206-1207, 'S', 1208-1209, 'S', 1210-1211, 'S', 1212-1213, 'S', 1214-1215, 'S', 1216-1217, 'S', 1218-1219, 'S', 1220-1221, 'S', 1222-1223, 'S', 1224-1225, 'S', 1226-1227, 'S', 1228-1229, 'S', 1230-1231, 'S', 1232-1233, 'S', 1234-1235, 'S', 1236-1237, 'S', 1238-1239, 'S', 1240-1241, 'S', 1242-1243, 'S', 1244-1245, 'S', 1246-1247, 'S', 1248-1249, 'S', 1250-1251, 'S', 1252-1253, 'S', 1254-1255, 'S', 1256-1257, 'S', 1258-1259, 'S', 1260-1261, 'S', 1262-1263, 'S', 1264-1265, 'S', 1266-1267, 'S', 1268-1269, 'S', 1270-1271, 'S', 1272-1273, 'S', 1274-1275, 'S', 1276-1277, 'S', 1278-1279, 'S', 1280-1281, 'S', 1282-1283, 'S', 1284-1285, 'S', 1286-1287, 'S', 1288-1289, 'S', 1290-1291, 'S', 1292-1293, 'S', 1294-1295, 'S', 1296-1297, 'S', 1298-1299, 'S', 1300-1301, 'S', 1302-1303, 'S', 1304-1305, 'S', 1306-1307, 'S', 1308-1309, 'S', 1310-1311, 'S', 1312-1313, 'S', 1314-1315, 'S', 1316-1317, 'S', 1318-1319, 'S', 1320-1321, 'S', 1322-1323, 'S', 1324-1325, 'S', 1326-1327, 'S', 1328-1329, 'S', 1330-1331, 'S', 1332-1333, 'S', 1334-1335, 'S', 1336-1337, 'S', 1338-1339, 'S', 1340-1341, 'S', 1342-1343, 'S', 1344-1345, 'S', 1346-1347, 'S', 1348-1349, 'S', 1350-1351, 'S', 1352-1353, 'S', 1354-1355, 'S', 1356-1357, 'S', 1358-1359, 'S', 1360-1361, 'S', 1362-1363, 'S', 1364-1365, 'S', 1366-1367, 'S', 1368-1369, 'S', 1370-1371, 'S', 1372-1373, 'S', 1374-1375, 'S', 1376-1377, 'S', 1378-1379, 'S', 1380-1381, 'S', 1382-1383, 'S', 1384-1385, 'S', 1386-1387, 'S', 1388-1389, 'S', 1390-1391, 'S', 1392-1393, 'S', 1394-1395, 'S', 1396-1397, 'S', 1398-1399, 'S', 1400-1401, 'S', 1402-1403, 'S', 1404-1405, 'S', 1406-1407, 'S', 1408-1409, 'S', 1410-1411, 'S', 1412-1413, 'S', 1414-1415, 'S', 1416-1417, 'S', 1418-1419, 'S', 1420-1421, 'S', 1422-1423, 'S', 1424-1425, 'S', 1426-1427, 'S', 1428-1429, 'S', 1430-1431, 'S', 1432-1433, 'S', 1434-1435, 'S', 1436-1437, 'S', 1438-1439, 'S', 1440-1441, 'S', 1442-1443, 'S', 1444-1445, 'S', 1446-1447, 'S', 1448-1449, 'S', 1450-1451, 'S', 1452-1453, 'S', 1454-1455, 'S', 1456-1457, 'S', 1458-1459, 'S', 1460-1461, 'S', 1462-1463, 'S', 1464-1465, 'S', 1466-1467, 'S', 1468-1469, 'S', 1470-1471, 'S', 1472-1473, 'S', 1474-1475, 'S', 1476-1477, 'S', 1478-1479, 'S', 1480-1481, 'S', 1482-1483, 'S', 1484-1485, 'S', 1486-1487, 'S', 1488-1489, 'S', 1490-1491, 'S', 1492-1493, 'S', 1494-1495, 'S', 1496-1497, 'S', 1498-1499, 'S', 1500-1501, 'S', 1502-1503, 'S', 1504-1505, 'S', 1506-1507, 'S', 1508-1509, 'S', 1510-1511, 'S', 1512-1513, 'S', 1514-1515, 'S', 1516-1517, 'S', 1518-1519, 'S', 1520-1521, 'S', 1522-1523, 'S', 1524-1525, 'S', 1526-1527, 'S', 1528-1529, 'S', 1530-1531, 'S', 1532-1533, 'S', 1534-1535, 'S', 1536-1537, 'S', 1538-1539, 'S', 1540-1541, 'S', 1542-1543, 'S', 1544-1545, 'S', 1546-1547, 'S', 1548-1549, 'S', 1550-1551, 'S', 1552-1553, 'S', 1554-1555, 'S', 1556-1557, 'S', 1558-1559, 'S', 1560-1561, 'S', 1562-1563, 'S', 1564-1565, 'S', 1566-1567, 'S', 1568-1569, 'S', 1570-1571, 'S', 1572-1573, 'S', 1574-1575, 'S', 1576-1577, 'S', 1578-1579, 'S', 1580-1581, 'S', 1582-1583, 'S', 1584-1585, 'S', 1586-1587, 'S', 1588-1589, 'S', 1590-1591, 'S', 1592-1593, 'S', 1594-1595, 'S', 1596-1597, 'S', 1598-1599, 'S', 1600-1601, 'S', 1602-1603, 'S', 1604-1605, 'S', 1606-1607, 'S', 1608-1609, 'S', 1610-1611, 'S', 1612-1613, 'S', 1614-1615, 'S', 1616-1617, 'S', 1618-1619, 'S', 1620-1621, 'S', 1622-1623, 'S', 1624-1625, 'S', 1626-1627, 'S', 1628-1629, 'S', 1630-1631, 'S', 1632-1633, 'S', 1634-1635, 'S', 1636-1637, 'S', 1638-1639, 'S', 1640-1641, 'S', 1642-1643, 'S', 1644-1645, 'S', 1646-1647, 'S', 1648-1649, 'S', 1650-1651, 'S', 1652-1653, 'S', 1654-1655, 'S', 1656-1657, 'S', 1658-1659, 'S', 1660-1661, 'S', 1662-1663, 'S', 1664-1665, 'S', 1666-1667, 'S', 1668-1669, 'S', 1670-1671, 'S', 1672-1673, 'S', 1674-1675, 'S', 1676-1677, 'S', 1678-1679, 'S', 1680-1681, 'S', 1682-1683, 'S', 1684-1685, 'S', 1686-1687, 'S', 1688-1689, 'S', 1690-1691, 'S', 1692-1693, 'S', 1694-1695, 'S', 1696-1697, 'S', 1698-1699, 'S', 1700-1701, 'S', 1702-1703, 'S', 1704-1705, 'S', 1706-1707, 'S', 1708-1709, 'S', 1710-1711, 'S', 1712-1713, 'S', 1714-1715, 'S', 1716-1717, 'S', 1718-1719, 'S', 1720-1721, 'S', 1722-1723, 'S', 1724-1725, 'S', 1726-1727, 'S', 1728-1729, 'S', 1730-1731, 'S', 1732-1733, 'S', 1734-1735, 'S', 1736-1737, 'S', 1738-1739, 'S', 1740-1741, 'S', 1742-1743, 'S', 1744-1745, 'S', 1746-1747, 'S', 1748-1749, 'S', 1750-1751, 'S', 1752-1753, 'S', 1754-1755, 'S', 1756-1757, 'S', 1758-1759, 'S', 1760-1761, 'S', 1762-1763, 'S', 1764-1765, 'S', 1766-1767, 'S', 1768-1769, 'S', 1770-1771, 'S', 1772-1773, 'S', 1774-1775, 'S', 1776-1777, 'S', 1778-1779, 'S', 1780-1781, 'S', 1782-1783, 'S', 1784-1785, 'S', 1786-1787, 'S', 1788-1789, 'S', 1790-1791, 'S', 1792-1793, 'S', 1794-1795, 'S', 1796-1797, 'S', 1798-1799, 'S', 1800-1801, 'S', 1802-1803, 'S', 1804-1805, 'S', 1806-1807, 'S', 1808-1809, 'S', 1810-1811, 'S', 1812-1813, 'S', 1814-1815, 'S', 1816-1817, 'S', 1818-1819, 'S', 1820-1821, 'S', 1822-1823, 'S', 1824-1825, 'S', 1826-1827, 'S', 1828-1829, 'S', 1830-1831, 'S', 1832-1833, 'S', 1834-1835, 'S', 1836-1837, 'S', 1838-1839, 'S', 1840-1841, 'S', 1842-1843, 'S', 1844-1845, 'S', 1846-1847, 'S', 1848-1849, 'S', 1850-1851, 'S', 1852-1853, 'S', 1854-1855, 'S', 1856-1857, 'S', 1858-1859, 'S', 1860-1861, 'S', 1862-1863, 'S', 1864-1865, 'S', 1866-1867, 'S', 1868-1869, 'S', 1870-1871, 'S', 1872-1873, 'S', 1874-1875, 'S', 1876-1877, 'S', 1878-1879, 'S', 1880-1881, 'S', 1882-1883, 'S', 1884-1885, 'S', 1886-1887, 'S', 1888-1889, 'S', 1890-1891, 'S', 1892-1893, 'S', 1894-1895, 'S', 1896-1897, 'S', 1898-1899, 'S', 1900-1901, 'S', 1902-1903, 'S', 1904-1905, 'S', 1906-1907, 'S', 1908-1909, 'S', 1910-1911, 'S', 1912-1913, 'S', 1914-1915, 'S', 1916-1917, 'S', 1918-1919, 'S', 1920-1921, 'S', 1922-1923, 'S', 1924-1925, 'S', 1926-1927, 'S', 1928-1929, 'S', 1930-1931, 'S', 1932-1933, 'S', 1934-1935, 'S', 1936-1937, 'S', 1938-1939, 'S', 1940-1941, 'S', 1942-1943, 'S', 1944-1945, 'S', 1946-1947, 'S', 1948-1949, 'S', 1950-1951, 'S', 1952-1953, 'S', 1954-1955, 'S', 1956-1957, 'S', 1958-1959, 'S', 1960-1961, 'S', 1962-1963, 'S', 1964-1965, 'S', 1966-1967, 'S', 1968-1969, 'S', 1970-1971, 'S', 1972-1973, 'S', 1974-1975, 'S', 1976-1977, 'S', 1978-1979, 'S', 1980-1981, 'S', 1982-1983, 'S', 1984-1985, 'S', 1986-1987, 'S', 1988-1989, 'S', 1990-1991, 'S', 1992-1993, 'S', 1994-1995, 'S', 1996-1997, 'S', 1998-1999, 'S', 2000-2001, 'S', 2002-2003, 'S', 2004-2005, 'S', 2006-2007, 'S', 2008-2009, 'S', 2010-2011, 'S', 2012-2013, 'S', 2014-2015, 'S', 2016-2017, 'S', 2018-2019, 'S', 2020-2021, 'S', 2022-2023, 'S', 2024-2025, 'S', 2026-2027, 'S', 2028-2029, 'S', 2030-2031, 'S', 2032-2033, 'S', 2034-2035, 'S', 2036-2037, 'S', 2038-2039, 'S', 2040-2041, 'S', 2042-2043, 'S', 2044-2045, 'S', 2046-2047, 'S', 2048-2049, 'S', 2050-2051, 'S', 2052-2053, 'S', 2054-2055, 'S', 2056-2057, 'S', 2058-2059, 'S', 2060-2061, 'S', 2062-2063, 'S', 2064-2065, 'S', 2066-2067, 'S', 2068-2069, 'S', 2070-2071, 'S', 2072-2073, 'S', 2074-2075, 'S', 2076-2077, 'S', 2078-2079, 'S', 2080-2081, 'S', 2082-2083, 'S', 2084-2085, 'S', 2086-2087, 'S', 2088-2089, 'S', 2090-2091, 'S', 2092-2093, 'S', 2094-2095, 'S', 2096-2097, 'S', 2098-2099, 'S', 2100-2101, 'S', 2102-2103, 'S', 2104-2105, 'S', 2106-2107, 'S', 2108-2109, 'S', 2110-2111, 'S', 2112-2113, 'S', 2114-2115, 'S', 2116-2117, 'S', 2118-2119, 'S', 2120-2121, 'S', 2122-2123, 'S', 2124-2125, 'S', 2126-2127, 'S', 2128-2129, 'S', 2130-2131, 'S', 2132-2133, 'S', 2134-2135, 'S', 2136-2137, 'S', 2138-2139, 'S', 2140-2141, 'S', 2142-2143, 'S', 2144-2145, 'S', 2146-2147, 'S', 2148-2149, 'S', 2150-2151, 'S', 2152-2153, 'S', 2154-2155, 'S', 2156-2157, 'S', 2158-2159, 'S', 2160-2161, 'S', 2162-2163, 'S', 2164-2165, 'S', 2166-2167, 'S', 2168-2169, 'S', 2170-2171, 'S', 2172-2173, 'S', 2174-2175, 'S', 2176-2177, 'S', 2178-2179, 'S', 2180-2181, 'S', 2182-2183, 'S', 2184-2185, 'S', 2186-2187, 'S', 2188-2189, 'S', 2190-2191, 'S', 2192-2193, 'S', 2194-2195, 'S', 2196-2197, 'S', 2198-2199, 'S', 2200-2201, 'S', 2202-2203, 'S', 2204-2205, 'S', 2206-2207, 'S', 2208-2209, 'S', 2210-2211, 'S', 2212-2213, 'S', 2214-2215, 'S', 2216-2217, 'S', 2218-2219, 'S', 2220-2221, 'S', 2222-2223, 'S', 2224-2225, 'S', 2226-2227, 'S', 2228-2229, 'S', 2230-2231, 'S', 2232-2233, 'S', 2234-2235, 'S', 2236-2237, 'S', 2238-2239, 'S', 2240-2241, 'S', 2242-2243, 'S', 2244-2245, 'S', 2246-2247, 'S', 2248-2249, 'S', 2250-2251, 'S', 2252-2253, 'S', 2254-2255, 'S', 2256-2257, 'S', 2258-2259, 'S', 2260-2261, 'S', 2262-2263, 'S', 2264-2265, 'S', 2266-2267, 'S', 2268-2269, 'S', 2270-2271, 'S', 2272-2273, 'S', 2274-2275, 'S', 2276-2277, 'S', 2278-2279, 'S', 2280-2281, 'S', 2282-2283, 'S', 2284-2285, 'S', 2286-2287, 'S', 2288-2289, 'S', 2290-2291, 'S', 2292-2293, 'S', 2294-2295, 'S', 2296-2297, 'S', 2298-2299, 'S', 2300-2301, 'S', 2302-2303, 'S', 2304-2305, 'S', 2306-2307, 'S', 2308-2309, 'S', 2310-2311, 'S', 2312-2313, 'S', 2314-2315, 'S', 2316-2317, 'S', 2318-2319, 'S', 2320-2321, 'S', 2322-2323, 'S', 2324-2325, 'S', 2326-2327, 'S', 2328-2329, 'S', 2330-2331, 'S', 2332-2333, 'S', 2334-2335, 'S', 2336-2337, 'S', 2338-2339, 'S', 2340-2341, 'S', 2342-2343, 'S', 2344-2345, 'S', 2346-2347, 'S', 2348-2349, 'S', 2350-2351, 'S', 2352-2353, 'S', 2354-2355, 'S', 2356-2357, 'S', 2358-2359, 'S', 2360-2361, 'S', 2362-2363, 'S', 2364-2365, 'S', 2366-2367, 'S', 2368-2369, 'S', 2370-2371, 'S', 2372-2373, 'S', 2374-2375, 'S', 2376-2377, 'S', 2378-2379, 'S', 2380-2381, 'S', 2382-2383, 'S', 2384-2385, 'S', 2386-2387, 'S', 2388-2389, 'S', 2390-2391, 'S', 2392-2393, 'S', 2394-2395, 'S', 2396-2397, 'S', 2398-2399, 'S', 2400-2401, 'S', 2402-2403, 'S', 2404-2405, 'S', 2406-2407, 'S', 2408-2409, 'S', 2410-2411, 'S', 2412-2413, 'S', 2414-2415, 'S', 2416-2417, 'S', 2418-2419, 'S', 2420-2421, 'S', 2422-2423, 'S', 2424-2425, 'S', 2426-2427, 'S', 2428-2429, 'S', 2430-2431, 'S', 2432-2433, 'S', 2434-2435, 'S', 2436-2437, 'S', 2438-2439, 'S', 2440-2441, 'S', 2442-2443, 'S', 2444-2445, 'S', 2446-2447, 'S', 2448-2449, 'S', 2450-2451, 'S', 2452-2453, 'S', 2454-2455, 'S', 2456-2457, 'S', 2458-2459, 'S', 2460-2461, 'S', 2462-2463, 'S', 2464-2465, 'S', 2466-2467, 'S', 2468-2469, 'S', 2470-2471, 'S', 2472-2473, 'S', 2474-2475, 'S', 2476-2477, 'S', 2478-2479, 'S', 2480-2481, 'S', 2482-2483, 'S', 2484-2485, 'S', 2486-2487, 'S', 2488-2489, 'S', 2490-2491, 'S', 2492-2493, 'S', 2494-2495, 'S', 2496-2497, 'S', 2498-2499, 'S', 2500-2501, 'S', 2502-2503, 'S', 2504-2505, 'S', 2506-2507, 'S', 2508-2509, 'S', 2510-2511, 'S', 2512-2513, 'S', 2514-2515, 'S', 2516-2517, 'S', 2518-2519, 'S', 2520-2521, 'S', 2522-2523, 'S', 2524-2525, 'S', 2526-2527, 'S', 2528-2529, 'S', 2530-2531, 'S', 2532-2533, 'S', 2534-2535, 'S', 2536-2537, 'S', 2538-2539, 'S', 2540-2541, 'S', 2542-2543, 'S', 2544-2545, 'S', 2546-2547, 'S', 2548-2549, 'S', 2550-2551, 'S', 2552-2553, 'S', 2554-2555, 'S', 2556-2557, 'S', 2558-2559, 'S', 2560-2561, 'S', 2562-2563, 'S', 2564-2565, 'S', 2566-2567, 'S', 2568-2569, 'S', 2570-2571, 'S', 2572-2573, 'S', 2574-2575, 'S', 2576-2577, 'S', 2578-2579, 'S', 2580-2581, 'S', 2582-2583, 'S', 2584-2585, 'S', 2586-2587, 'S', 2588-2589, 'S', 2590-2591, 'S', 2592-2593, 'S', 2594-2595, 'S', 2596-2597, 'S', 2598-2599, 'S', 2600-2601, 'S', 2602-2603, 'S', 2604-2605, 'S', 2606-2607, 'S', 2608-2609, 'S', 2610-2611, 'S', 2612-2613, 'S', 2614-2615, 'S', 2616-2617, 'S', 2618-2619, 'S', 2620-2621, 'S', 2622-2623, 'S', 2624-2625, 'S', 2626-2627, 'S', 2628-2629, 'S', 2630-2631, 'S', 2632-2633, 'S', 2634-2635, 'S', 2636-2637, 'S', 2638-2639, 'S', 2640-2641, 'S', 2642-2643, 'S', 2644-2645, 'S', 2646-2647, 'S', 2648-2649, 'S', 2650-2651, 'S', 2652-2653, 'S', 2654-2655, 'S', 2656-2657, 'S', 2658-2659, 'S', 2660-2661, 'S', 2662-2663, 'S', 2664-2665, 'S', 2666-2667, 'S', 2668-2669, 'S', 2670-2671, 'S', 2672-2673, 'S', 2674-2675, 'S', 2676-2677, 'S', 2678-2679, 'S', 2680-2681, 'S', 2682-2683, 'S', 2684-2685, 'S', 2686-2687, 'S', 2688-2689, 'S', 2690-2691, 'S', 2692-2693, 'S', 2694-2695, 'S', 2696-2697, 'S', 2698-2699, 'S', 2700-2701, 'S', 2702-2703, 'S', 2704-2705, 'S', 2706-2707, 'S', 2708-2709, 'S', 2710-2711, 'S', 2712-2713, 'S', 2714-2715, 'S', 2716-2717, 'S', 2718-2719, 'S', 2720-2721, 'S', 2722-2723, 'S', 2724-2725, 'S', 2726-2727, 'S', 2728-2729, 'S', 2730-2731, 'S', 2732-2733, 'S', 2734-2735, 'S', 2736-2737, 'S', 2738-2739, 'S', 2740-2741, 'S', 2742-2743, 'S', 2744-2745, 'S', 2746-2747, 'S', 2748-2749, 'S', 2750-2751, 'S', 2752-2753, 'S', 2754-2755, 'S', 2756-2757, 'S', 2758-2759, 'S', 2760-2761, 'S', 2762-2763, 'S', 2764-2765, 'S', 2766-2767, 'S', 2768-2769, 'S', 2770-2771, 'S', 2772-2773, 'S', 2774-2775, 'S', 2776-2777, 'S', 2778-2779, 'S', 2780-2781, 'S', 2782-2783, 'S', 2784-2785, 'S', 2786-2787, 'S', 2788-2789, 'S', 2790-2791, 'S', 2792-2793, 'S', 2794-2795, 'S', 2796-2797, 'S', 2798-2799, 'S', 2800-2801, 'S', 2802-2803, 'S

A/Accession: A27720  
 A/Molecule type: mRNA  
 A/Residues: 1-1223 <LAX>  
 A/Cross-references: GB:M20386  
 R/Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rotman, F.M.; Citterenden, L.B.; Raines, M.  
 Cell 41, 719-726, 1985  
 A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro  
 A/Reference number: A00643; MUID:85228222; PMID:2988784  
 A/Accession: A00643  
 A/Molecule type: mRNA  
 A/Residues: 585-1223 <NIL>  
 A/Cross-references: GB:M10066  
 C/Genetics:  
 A:Gene:erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C/Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor  
 C/KeyWords: protein kinase  
 F:1-30/Domain: signal sequence #status predicted <Sig>  
 F:1-1223/Product: epidermal growth factor receptor #status predicted <MAT>  
 F:1-654/Domain: extracellular #status predicted <EXT>  
 F:31-307/Domain: EGF receptor extracellular domain repeat <EEL>  
 F:397-610/Domain: EGF receptor extracellular domain repeat <EE2>  
 F:655-677/Domain: transmembrane #status predicted <TMM>  
 F:678-1223/Domain: intracellular #status predicted <INT>  
 F:719-984/Domain: protein kinase homology <KIN>  
 F:727-735/Region: protein kinase ATP-binding motif  
 F:136-202,280,361,370,422,575,580,615,635/Binding site: carboxylate (Thr) (covalent) #s  
 F:192,650/Binding site: carboxylate (Ser) (covalent) #status predicted  
 F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:754/Active site: Lys #status predicted  
 F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.6%; Score 3049.5; DB 1; Length 1223;  
 Best Local Similarity 48.3%; Pred. No. 2,66-118;  
 Matches 626; Conservative 166; Mismatches 360; Indels 145; Gaps 25;

8 RMCILLALPPGAA-----STQVCTGTMKRLPASPETHLDMRLHYCGQVQGNLE 61  
 13 RGAVALVLLGLGALCSAVEBEKKVCGGTNNKLTQGHVDHFTSLQRMNNECVLSNLE 72  
 62 LTYLPNNAFLSFDQDQVQGYVLAHNOYRQVPLRLIIVGTQIFEDNALAVLDNGD 121  
 73 IYVENRDLTFLKTOEVAQYVLAHNVDAIPLNIOIIRGNVLYXNSFALAVLSNTH 132  
 122 PLNNTPTVPGASPGSLREQLRSLEILKGVLIQNPOLQODITIMWIDHKNQAL 181  
 133 -NNKTO-----GLRELPKRLISILNGSVKISNNPKLCNMDVLMNDIIDTSRK-PL 182  
 182 TLID-TNRSRACHPSCGKSGSRCWGESSEDCQSILTRVCAGCA-RCKGPLPTDCHEQ 239  
 183 TVLDFASNLSCPKCHPNCCTEDHCWAGGQNCQITLTKVCAQCGSGRCGRKVPSCCHNQ 242  
 240 CAAGCTGPKGSDCLCLHNHSGICELHCPALVYNTDFESMPNPERGRTGASCYTAC 299  
 243 CAAGCTGPKGSDCLCLCRKFRDATCCTCPPLVLTNPYQNDVPEGYSFGATCYEC 302  
 300 PNYLSTDVSGCTLVCPLHNOEYTAEDGTQCEKSKPCARVCYLGMQYIANKSKFGI 359  
 303 PHNYVYTDHSCVRSQNTDTEYV-BENGVKCKCKDGLCSKCNQNIIGELKGLS-INA 360  
 360 TLEL-FAGCKKIFGSLAPLPESPDGASNTAPLOEQIQLVETLEITTYIYISAMPDS 418  
 361 TUIDSPKNTKINGVSLIPVAFIDAFKTLPLDPKLDVRYTKEISGFLLOQAMDN 420  
 419 LPDLVFOQLQVIRIRLHNGAVSLTLOGLISMTGLSLRELSGLLIHNNTHLCVH 478  
 421 ARDVAPELLEIRRTKHQGOYSLAVNLKISGLASLKEISGDIAMKNKYLCTAD 480  
 479 TVPMDLPNNPQALLHNPREDCEVGBGLACHOLCARGHGWPFGPTQVCNCSQGLRGQ 538  
 481 TMMKSLFATOSQKTKIIONENKNDCTADRHVCDPLGSDVGWGPFGPHCSRFPSRQK 540  
 539 ECVEECRVLQGLPFREYVARHCLPCHPEQONG---SVTGGPEADCCVACAHYKDPF 595

541 ECVKQCNILQGEPRFEREDSKLPCSHSCLVONSTAYNTTSCSPGPDHCKMCAHPIIDGP 600  
 596 CVARPGSVKPDLSWPMKFPDEGACQOPPEINTHSCVLDLGGCAEGRASPLTSTIF 655  
 601 CVKAPPAVLGENDTL-VKTDANAVQLCHPNTROCKPGLEGCP--NGSTTFPIA 656  
 656 NNFTVFWLRYPKVSASHLEKROQKIRKTYMRRLLOETLEVEPTPSGAMPNQAMRIL 715  
 657 AGVGGGLCLVVGIGLILYRRRIYAKRLTLRLLQRELEVEPTPSGAMPNQAMRIL 716  
 716 KETELRKXVVGSGAFGVYKGIWPDGNNKIPVATVLRBNTSPKANKELDEAYMA 775  
 717 KETEFKXKVGSGAFGVYKGIWPDGNNKIPVATVLRBNTSPKANKELDEAYMA 776  
 776 GVSRYVRLIGICTSTVQVLTQMPYGLCLDHRBGRGLSDLLNMCQIAKMSY 835  
 777 SVDNHVRRLIGICTSTVQVLTQMPYGLCLDHRBGRGLSDLLNMCQIAKMSY 836  
 836 LEDVALVRDLAARNVLYKSPBNHYKIDFGRLALDIDETEVHADGKVPKIMWALLESIL 895  
 837 LEERLVARDLAARNVLYKSPBNHYKIDFGRLALDIDETEVHADGKVPKIMWALLESIL 896  
 896 RRRFTHSDVMSYGVTVWELMTFGAKPFDGIIPAREIPDLLEKGERLPQPICTIDVYIM 955  
 897 HRIYHSDVMSYGVTVWELMTFGAKPFDGIIPAREIPDLLEKGERLPQPICTIDVYIM 956  
 956 VKCMWIDSECPREELVSEFSRMAPRQRFVQI-NEDLPASPLDSTFRSLDEDDM 1014  
 957 VKCMWIDSECPREELVSEFSRMAPRQRFVQI-NEDLPASPLDSTFRSLDEDDM 1016  
 1015 GDLYDAEYLYPQCGFPCPDPAAGAKGVHHRSSSTRSGGDLTGLEPSEBAPSP 1074  
 1017 EDYDAEYLYPQCGFPCPDPAAGAKGVHHRSSSTRSGGDLTGLEPSEBAPSP 1074  
 1075 L-----APSEGAGDVDFGDLGMAAGLQSLPHDPSPLQRYSEDETPVLPST--DGY 1127  
 1043 LLSLSLSTSNNSATNCID-----RNGQCHPREDSEFVQRYSSDPGFLNLESDGF 1094  
 1128 VAPLTPSPQEPYVNPDPVPPSPRPGPLPAPAPATLERAKTLSPGXGVKVDVF-- 1185  
 1095 L-----PAPEYVQ--LMPKPS-----TAMVQNIYNNISLT 1125  
 1186 -----AFGAVENPEYLTQGGAAAPRPPAPSPAFDNLVYWDQ----- 1225  
 1126 AISKLPMSRYQNSHTAVDNPEYL-----NTNOSPAAKTAFESSPWYIOGNHQN 1177  
 1226 -DPE-----RGAPSTFKGTAEENPEYLGDPV 1254  
 1178 LDNPDYQDDPLPNETKENGLLKVALENPEYLRVAP 1214

RESULT 7  
 A47253  
 epidermal growth factor receptor, HER4 - human  
 C/Species: Homo sapiens (man)  
 C/Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999  
 C/Accession: A47253  
 R/Plowman, G.D.; Culicson, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993  
 A/Title: Ligand-specific activation of HER4/Plowman, a fourth member of the epider  
 A/Reference number: A47253; MUID:93189574; PMID:8333326  
 A/Accession: A47253  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: nucleic acid  
 A/Residues: 1-1308 <ELO>  
 A/Cross-references: GB:L07868; NID:9337359; PIDN:AAB59446.1; PID:9337360  
 A/Note: sequence extracted from NCBI backbone (NCBI:P126842)  
 C/Suprafamily: epidermal growth factor receptor; protein kinase homology  
 C/Keywords: ATP; growth factor receptor  
 F:716-981/Domain: protein kinase ATP-binding motif  
 F:724-732/Region: protein kinase ATP-binding motif

```

Query Match      42.7%; Score 2920.5; DB 2; Length 1308;
Best Local Similarity 44.6%; Pred. No. 5.5e-113;
Matches 604; Conservative 178; Mismatches 384; Indels 187; Gaps 26;

QY 9 MGLLLALLPAGAA-----STQVCTGDMKRLPASPETHLDMRLHLVQGCQVQGNLELY 64
DB 8 MWWWLLVANGVYQPSDQSVCACTENKLSLSLDEQYRALRYENEVEVMGNLEIS 67
QY 65 LPTNASLFLQDIOEVQGYLLAHNOVQVPLORLIRVGTOLFEDNYALAVINDGPLYN 124
DB 68 IEHNDLSFLRSYREVYGYVALNQFRYLPENRIIRGTXYERVALAFYLRKDG 127
QY 125 NTPVTGASPGRLREQLRSLEILKGVLIQNPOLCYODTILMDIETHKNNQALTLI 184
DB 128 NF-----GLQELKLVTEILNGVYVQNKLCYADTIHMDIYRNPSPSLTIV 178
QY 185 DTRSPACHPSCPMKSGRCWGSSEDCSLTFTVACAGC-ARCKPPLPTDCEHCQAG 243
DB 179 STNGSGGCRCHKSCG-RCWGPTEHNCQTLTRVCAEQDCRCYPRVSDCCHREGAG 237
QY 244 CTGPKSDCLACHFNHSGICELHCPALVYNTDESMNPREGRTFGASCYTACPVNY 303
DB 238 CSGKPDTCFACMNFDSGACVTCQCFYVYNTTQLEHNEAKTYGAFVCKCPHNF 297
QY 304 LSTDVGSCTLVCEPLHNOEVTABDQRCCKSKPCARVCYGLMOKYIKANSKFTIGTELE 363
DB 298 V-VDSSSCVACPSKMEV-BENGDKCKPCTDIPCACGIGTSGLSMAQTVDSNIDX 355
QY 364 PACGKTFGLATLPESFDDPASNTPLOPQOLVETLEITGYLYISAMPDLSPLDS 423
DB 356 FINCTKINGNLIPLVTHGDYNAIEADPEKLVFRYREITGFLNTQSPMPTDPS 415
QY 424 VEQNLVIRGRILHNGAYSLTLOGLISWLGRLSRLPSGLALIHNTHLCEVAFVPM 483
DB 416 VFSNLVITGRVLYSGSLILKQGTSLQFOSLKSIGNYITDINSCLCYHINMT 475
QY 484 QLEFNPHQALLHTANPEDECVGEGLACHQLCARGHCWGPFPQVNCQSLFEGQVEE 543
DB 476 TLFSTINQIRIVIRNRAKENTAGMVCNHLCSDDGCGWGPQCCSFRFSGRICIES 535
QY 544 CRVYQGLPREVYVARHCLPCHPEGOP-QNGSVTCFGEADQCAVACHYKPPRCVAPCS 602
DB 536 CNLDGFRFENOSICECDPOCKMEDGLITCHGFPNCTKCHFKGCPVCEKCPD 585
QY 603 GVKEDELSYMPWKEPDEGAQCOPFINCTHSCVDLDKGPABORASPLTSLP-----NNF 658
DB 596 GLQGANSP--IFKYADPDRECHPCHPCTOGCNGPTSHDC-----IYPTWGHG 642
QY 659 TVSFMLRVPKYSAS-----HLEKROCKIKRYTMRLLOTELVETLPS 703
DB 643 TLPGHARTPLIAAGVIGLFLVIVGLTFPAVYRRKSIKKRALRRL-ETELVEPLTPS 701
QY 704 GAMPNOQMRILKTELKRYKVLGSGAFYVYKGIWIPDENVKIPAIKVLRENTSPKA 763
DB 702 GTANQOQRLKTELKRYKVLGSGAFYVYKGIWIPDENVKIPAIKVLRENTSPKA 761
QY 764 NKEILDEAVYVAGSGPYVSRLLGICITSTVQVLTQMPYGLCLDHYREKRGKLSODLL 823
DB 762 NVEFMDEALIMASMDHPLVRLGVCSPITQVLTQMPYGLCLDHYREKRGKLSODLL 821
QY 824 NMCOQIAKMSYLEDVLRHDLAARNVLYKSPHVKITPTFGALRLDIDETEYHADGK 883
DB 822 NMCOQIAKMSYLEDVLRHDLAARNVLYKSPHVKITPTFGALRLDIDETEYHADGK 881
QY 884 VPIKMALESILRRFTHQSDVSYGVYVWELMTFGAKPYDGPAREIPLLEKGERLPQ 943
DB 882 MPIKMALECIHYKFFTHQSDVSYGVYVWELMTFGAKPYDGPAREIPLLEKGERLPQ 941
QY 944 PICTIVYVIMVCMWIMDSRCRPRFELVSEFSRMAKDPORFVIONED-LGPASPLDS 1002
DB 942 PICTIVYVIMVCMWIMDSRCRPRFELVSEFSRMAKDPORFVIONED-LGPASPLDS 1001
QY 1003 TFRSLLEDMDGLVDAEYLVQGFPCDPAPAGAGVHRRHRSSTRSGGDLTLG 1062

RESULTS
6306142
protein-tyrosine kinase (EC 2.7.1.112) mrk-Y precursor - southern platyfish
N.Alternate names: epidermal growth factor receptor homolog; kinase-related transfo
C.Species: Xiphophorus maculatus (southern platyfish)
C.Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C.Accession: S06142; S13809
C.Mitochondrion: J.; Adam, D.; Maltischek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Ro
Nature 341, 415-421, 1989
A.Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing T
A.Reference number: S06142; MUID:90015140; PMID:2791166
A.Accession: S06142
A.Molecule type: DNA
A.Residues: 1-1166 <WTT>
A.Cross-references: EMBL:X16891; NID:G65290; PIDN:CAA34770.1; PID:G65291
R.Adam, D.; Maeueller, W.; Scharf, M.
C.Orgene 6, 73-80, 1991
A.Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xipho
A.Reference number: S13807; MUID:91125882; PMID:1846957
A.Accession: S13809
A.Status: preliminary; translation not shown
A.Molecule type: DNA
A.Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A.Cross-references: EMBL:X56319; NID:G65284; PIDN:CAA39763.1; PID:G65285
C.Genetics:
A.Gene: mrk
A.Map position: Y
A.Introns: 872/3; 988/1; 947/1; 979/3; 1025/3; 1056/1
C.Superfamily: epidermal growth factor receptor; protein kinase homology
C.Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein;
F1-25/Domain: signal sequence #status predicted <SIG>
F1-26/166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F1-707-972/Domain: protein kinase homology <KIN>
F1-715-723/Region: protein kinase ATP-binding motif

Query Match      38.4%; Score 2627; DB 1; Length 1166;
Best Local Similarity 44.7%; Pred. No. 5.7e-101;
Matches 568; Conservative 165; Mismatches 393; Indels 144; Gaps 31;

QY 4 AALCRGGLLALLPAGAAST---OVCSTGDMKRLPASPETHLDMRLHLVQGCQVQGN 59
DB 8 AALLO--LLLVISISRCSTDPDRVCQGTSTQNM---LDNHYLRKMKVYSGCVAVLEN 62
QY 60 LELTYLPTNASLSPFDIOEVQGYLLAHNOVQVPLORLIRVGTOLFEDNYALAVIND 119
DB 63 LELTYLPTNASLSPFDIOEVQGYLLAHNOVQVPLORLIRVGTOLFEDNYALAVIND 112
QY 120 GDLNNTPTVTGASPGRLREQLRSLEILKGVLIQNPOLCYODTILMDIETHKNNQ 179

```

Db 123 YOK-NDPSP--DVTQVGLKQJQLSLNLTITLGGYKVSNNPLLCVETINMDIDVKTSP 179  
 QY 180 ALTLIDTRSRACHPCSPCKGSRMGSSSEDCSLTRTVAGGC-ARCKGPLPTDCHE 238  
 Db 180 TMLNLPFAFERQCCQKCHGCVNGSCWAPGHCQFTLLCAEQCNRRCKRPKIDCNE 239  
 QY 229 QCAAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTESNPNRGRYTPGASCTA 298  
 Db 240 HCAGGCTGPRAITDCLACDFNDGCKDTCPPKTYDVSQVVDNPNKITYFGAACVKE 299  
 QY 299 CPYNYLSTDVSGCTVCPHLNQEVAEDGTQCEKSKPCARVCVGLMOYIKANSKFIG 358  
 Db 300 CPNSVYVTE-GACVRSACAGLBYD-ENGKSKCPKCDQVCKCDGIGISGL-SNTIAVN 356  
 QY 359 ITEL-EPAGCKKIFGSLAFLPSPGDPASNTAPLOPQLOVFETLEITGYLYISAMPD 417  
 Db 357 STNIRSFENCKTINDIILNRRNSFEGDPHYKIGTMDPBNLNTLVVKEITGYLYIWMPE 416  
 QY 418 SLPLSTVFONTQVIRGRILNNGAYS-LTLQGLG:SWLGLRSLRELGSGLALIHNTLCLF 476  
 Db 417 NMTLSLVQNLLEIRGRITTFRSGFSFVAVQVRHQLQGLRSLKESVSGNVILKNTLQRLY 476  
 QY 477 VHTVPMQDLFRNPHQALHTANREDECEVGEGLACHQ:CARGHCMGPGPTQCVNCSQFLR 536  
 Db 477 ANTIMWRRLFRSEDDQSIEDART-----ENQCNNECSEDCGM-PGPTWCVSCLHVD 528  
 QY 537 GQCEVEGRVULGLPREYVNAHCLPCHPRECQPONGSVTCGFPADQVACAKYKPPFC 596  
 Db 529 GGRCVASCNLLQGEFREAOVDGRQCQOECLVQDTSITCGPBPANCSKSAHFQDPQC 588  
 QY 597 VAPCPGVKPDLSWYPIKFPDEEGACQPCINCHSCVDDDDKGPAAE-QRAPLTSLI 654  
 Db 589 IPRCHGILGSDTL-TMKVADKXGQCPCHQNTQCGSGGLSGCGDVISHSILAVGL 647  
 QY 655 FNNFTSFMLRVPKVSASHLEKRPQOKIKRYTMRLIOETELVPLTPSGAMPQAMRI 714  
 Db 648 VSGLLITVALLIVV--LRRRRRIK-RKRTIRCLQEKELVPLTPSGQAPQALRI 703  
 QY 715 LKETELRVKVLGSGAGTGYKGIWIDGDEVNKPVAIKYLRNTSKAKELIDEAYVM 774  
 Db 704 LKETEFKDRVLGSGAGTGYKGLMNPDEGNIRIPVAIKYLRNTSKAKELIDEAYVM 763  
 QY 775 AGVGSFYVSLGLCLSTVQVLQVLTQMPYGLDHYENRGLSQDLNWMQIAKMS 834  
 Db 764 ASVDHPHYCRLLGICLSAVQLVQLMPYGLDYVAQHORIQQGLNWMQIAKMS 823  
 QY 835 YLEDEVLVHRDLAARNLVKSPNVTITPGLANLIDETETHADGKVPYKMALESI 894  
 Db 824 YLEERHLVHRDLAARNLVKSPNVTITPGLANLIDETETHADGKVPYKMALESI 883  
 QY 895 LRRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICITIDYMI 954  
 Db 884 LQWYTHQSDVWSYGVTVWELMTFGSKPYDGIIPAKETASVLENGERLPPOPICITIEYMI 943  
 QY 955 MYKCMWIDSECRPFRELVSEFSMAADPQRFVVIQNEIDLGPASPLDSTFYRSLBEDDM 1014  
 Db 944 ILKCMWIDPSRRPFRELVSEFSQMAADPSRYLQI--NLPSLDRRLFSRLSSDD- 999  
 QY 1015 GDVDAEEYLVPQGGFCPPDPAPAGGMVHHRHSSSTRSGGGLITGLERSEEARSP 1074  
 Db 1000 -DAVDADBYLLPYKRI-----NRQGS-----EP 1021  
 QY 1075 LAPSEGAGSDVFDGLMGAKGLQSLPTHDPSPLOQYSEDPVY-PLPSETDYVAPLTC 1133  
 Db 1022 CIPPTGH-----PVRENSTILANISDPQNALMKLODH----- 1055  
 QY 1134 SPQREYVNOVDVRFQ-----PSPRE-----GPLP-AARPAATLTERAKTISPGKNGV 1180  
 Db 1056 ----EYVNOQSESTSSRLSDIYNNYEDLTDGWPVSLSQEAEITNSRREYVNTQNSL 1111  
 QY 1181 VKDVFAGGAVENFEYLTPOGGAAPCPHPAPASPAEDNLVYWDQDPREGAGAPSTFKGT 1240  
 Db 1112 ---PLVSSGSMDDPDY---QAG-----YQAAE-----LPQGLALTGGMFL 1146

QY 1241 PTANDEYLIG 1250  
 Db 1147 PAENLEYLG 1156  
 RESULT 9  
 A36223  
 kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 04-Oct-1991 #sequence, revision 13-Jan-1993 #ext\_change 17-Nov-2000  
 C:Accession: A36223, 159164  
 R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.  
 Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989  
 A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epide:  
 A:Accession number: A36223; MVID:90083234; PMID:2687875  
 A:Accession: A36223  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-1342 <KRA>  
 R:Plouman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todar:  
 Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990  
 A:Title: Molecular cloning and expression of another epidermal growth factor recept  
 A:Reference number: 159164; MVID:90311312; PMID:2164210  
 A:Accession: 159164  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>  
 A:Gene: GDB:ERBB3; HER3  
 A:Cross-references: GDB:119880; OMIM:190151  
 A:Map position: 12q13-12q13  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase  
 C:Keywords: ATP; phosphotransferase  
 F:707-972/Domain: protein kinase homology <Kin>  
 F:715-723/Region: protein kinase ATP-binding motif  
 Query Match 35.0%; Score 2389.5; DB 2; Length 1342;  
 Best Local Similarity 40.0%; Pred. No. 3,8e-91;  
 Matches 526; Conservative 192; Mismatches 459; Indels 139; Gaps 33;  
 QY 10 GLIALLPFGAA--SYQVCTGTDKRLRLASEPETHLDMRLHYQCGVQVQNLTYLPT 67  
 Db 11 GLIFSLARSGSEVNSQAVCPGLTNG:SVTGAENQYQITLYKLYEREVEVMGNLEIVTGH 70  
 QY 68 NASLSFLQDIOEVQGVV:FAHQVQVPLQRLRIYRGTLQFEDNYVALAVLDGDDPLNNTT 127  
 Db 71 NADLSFLQWIREVTGVVYVAMNFSITLPLNIRVVRGTQVYQKFAIFVM-----LNVNT 125  
 QY 128 PYTGASPGGLRELQRLSLTELKGGVLIQRNPQLCYODITLWKDIFHKNNQALTLIDTN 187  
 Db 126 ----NSSHALROLRLTQLTEILISGVYIEKNKLCGMHDIWRDIWRD---AEIVXD 178  
 QY 188 SARACPGCPMKGSRMGSSSEDCOSLTRTYCAGGC-ARCKGPLPTDCHEQCAAGCTG 246  
 Db 179 NGRSECPCHVECKG-RQWPGSEDCQTLTKTICAQQNGCHCGPNNQCHCECAGGCSG 237  
 QY 247 PKHSDCLACHFNHSGICELHCPALVTYNTDTESMPNREGRYTFGASCVTACPYNYLST 306  
 Db 238 PDDTCFACRHHNDGACVPRCPPLVYVKNKLLFQLEPNHTKYQYGVGVASCPHNFV-V 296  
 QY 307 DVGSGTIVPRLNQEYTAEDGTQCEKSKPCARVCYGLQMOYIRANSKF--IGITTELE- 363  
 Db 297 DQTSQVRACPDKMVD-KNGLKQCEPCQGLCPKACBGTG-----SGSRFQTVDSNIDG 350  
 QY 364 FAGCKKIFGSLAFLEESFQDPASNTAPLOPQLOVFETLEITGYLYISAMPDLSPLDS 423  
 Db 351 FVNCIKIQLNDLFLTLGNDGPMHKIPALDEKLVNFTVREITGYLYINQSPMKHNS 410  
 QY 424 VFQNTQVIRGRILNNGAYS-LTLQGLG:SWLGLRSLRELGSGLALIHNTLCLFVHTVPW 482

```

Db      411 VESNLTTIGRSLSYNRGFSLLIMKVLNVTSLGFRSLKEISAGRIYISANROLCVHSLNW 470
Qy      463 DOLFNPQALLHTA-NRPEDECVGEGLACHOLCARHGCMGRPPQCNQOFLRGCECV 541
Db      471 TKVLGPTPEERLDIKHNRPRRDCVAGKVCPLCSSGGCWGGPQCLISCRNYSRGVCV 530
Qy      542 EECRYLQGLPREYVNAHCLPCHPECPQNGSVTFCEGEADQCVCAHAKDPFCVACRP 601
Db      531 THCNFLNPEPRFAHAECSFCHPECCPMEGATNGSGSDTCAQCAHFRGPHVSSCP 590
Qy      602 SGVFPDLSYMPIMKRPDEBACQPCPINCTHSQVLDLXGCPAE-----QRASPLTGI 654
Db      591 HGVLCG-AKGPITYKPDVQNECRPCHENCTQCKGPELQDCLGQTLVLIGKTHLTMLAT 648
Qy      655 FNNFTVSPFLRVKVSASHLEKRRQOKIRKVTMRLLQETLVEPLTSGAMPQAOQRI 714
Db      649 IAGLVYIFMM-----IGTFTLYRGRRIQNKRAMRYLGRGESIEPLDS-EKANKVLARI 703
Qy      745 IKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRENTSFRANKELIDEAYVM 774
Db      704 FKETELRKVKVLGSGVGFVHKGVWIPGESIKIPVCIKVIDKSGRQSFQAVTDHMLAI 763
Qy      775 AGVSPVYSRLIGLITSTVOLVTOLMPYGLDHYRENRGRISGODLLNMCQIAKMS 834
Db      764 GSLDHAHIVRELGLCPGSSQLVLYPLGSLDHYRQHGALGQLLNMGVQALAKMY 823
Qy      835 YLEDELVYHRDLAANVLYKSPNHVKITDFGLARLLDIDETRYHADGGKVPDKMALESI 894
Db      824 YLEHGMVHRNLAAANVLLKSPQVADFGVADLPDDQQLLESAKTIPIKMALESI 883
Qy      895 LRRFTHOSDVMSYGVTTWELMTFGAKPYDGI PAEIPDLLEKGRLLQPPICITIDVYMI 954
Db      884 HFGKTYHOSDVMSYGVTTWELMTFGAEPYAGRLALEVPDLLEKGRLLQPPICITIDVYV 943
Qy      955 MYKCMNIDSECPREELVSEFPMARDPORFVVIQONDLGPA---SPLDSTFYRLIED 1011
Db      944 MYKCMNIDENIRPTKELANETRMARDPRYLVIKRS-GRGIAPGEHPHGLNKLKE 1002
Qy      1012 DDMGDLVDAEYLVPOQGFPCDPAPAGAGVHHHRSSSTRSGGGLTLGLER-SEEEA 1070
Db      1003 VELPELDDLDLLEED-----NLAITTLASALSLVGTILNRPBG 1043
Qy      1071 PPSPLAPSEGASDVFDGLGMAKGLQSLPTHD-PSPLQYSSDPVPP-----SE 1123
Db      1044 SGLSLSPSSGY-MPMNGNLGSCQESAVSSSEKCPRVSLH-----PMRGLASBS 1096
Qy      1124 TDGYVA-----PLTCSQPE---YVNPQVYRPPSPRSGP----- 1156
Db      1097 SEGHTGSEAEIQEKVMSCRSRSPRPRGDSAYHSQRHSLTPTVPLSPGLEEDV 1156
Qy      1157 ----LPAARPAATLEBAKTSP-GRNGV-----KDVAFAGAVENPEYILTPGGGAAP 1205
Db      1157 NGVMPDTHLKGTSSREGTSLSSVGLSYLGEEDD-----EEYENNRRRRSP 1208
Qy      1206 QHPPAPAPAFDNLVYMD-----QDPERGAPSTFGTPTAENPEYL 1249
Db      1209 -FHPRPSSLELGEYVMDVGSDSLASLGSTGSCGLHPVIMPRTAGTTPDDDYEM 1263

```

## RESULT 10

Jc4387  
 epidermal growth factor receptor homolog precursor - rat  
 N:Alternate names: ErbB3 protein, HER3 protein  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 17-Jan-1996 #sequence\_revision 19-Apr-1996 #text\_change 13-Nov-1998  
 C:Accession: Jc4387  
 R:Helixer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.  
 Gene 165, 273-284, 1995  
 A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.  
 A:Reference number: Jc4387; MUID:96096535; PMID:8522190  
 A:Accession: Jc4387  
 A:Molecule type: mRNA  
 A:Residues: 1-1339 <HEL>

```

A/Cross-reference: GB:U29339; NID:9515389; PID:9515390
A/Experimental source: liver
A/Note: The authors translated the codon AAC for residue 369 as Thr and GGT for res
A/Comment: This protein is a functional heregulin receptor that transduces signals
C/Genetics:
A/Genes: ErbB3
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinases )
C/Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prot
F:1-19/Domain: signal sequence #status predicted <Sig>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase homology <KIN>
F:773-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1219,1257,1259,1273,1286,1325/Binding site: phosphate (T

```

Query Match 33.7%; Score 2302.5; DB 2; Length 1339;  
 Best Local Similarity 40.3%; Pred. No. 148-87;  
 Matches 518; Conservative 169; Mismatches 439; Indels 159; Gaps 34;

```

Qy      3 LVALCRWGLLALPPGAA---STQVCTGDMRLRLPASPEHLDMLRHLVQCCOVQGN 59
Db      7 LQVLC-----FLSLAGSEMGNSQAVCPGTLNGLSTGADNQGTYQLTKYKCEVYMGN 62
Qy      60 LEITLYPTNASLSFLQDIQEVQGVYLIANQVQVPLQRLRYRGTQLEEDNALAVLDN 119
Db      63 LEITVLTGHNDLSFLQWIREVLYVYVANNRFEVPLRLNLRVVRGTQVYDGKRAIFVM-- 120
Qy      120 GDPINNTPTPTGASPGGLRELOLRSLTEILKGGVLLOBNPOLCYODTILMKDIFHKNNQL 179
Db      121 ----LNNTY-----NSHARLQKLTQLEILSGVYIEKNDLKHMTIDKRDIVRVR-- 170
Qy      180 ALTLIDITNRSRACHPCSPMKGSRCESESDQSLTRIVCAQC-ARCKGLPTDCHE 238
Db      171 GAETVYQNGANRCPCHVEVCKG-RWGPGRDDQILTKICAPQCRGCRGPRPNQCCHD 229
Qy      239 QCAAGCGPRGSCCLAHNHSIGIELHRLAVTYNTDFFESMPNDEGTYTGASCVTA 298
Db      230 ECAAGCGPRDTCFARRRNDGACPRRREPLVYNKLTFQLEPNHTLYQVGGCVAS 289
Qy      299 CPNYVLTSDVGSCTLVCPRLNQEVTAEQTQRCSEKSKPCARVYGL--GMQYIKANSKF 356
Db      290 CPNHFV-VDTFCVACRPDKMEVD-KHGLKMECPGGLCPKACEGTGSGSRQGYDSSN 347
Qy      357 IGITELFPACCKIKFQSLALPESFGCDPRASNTAPLOPELOQVFTLEITGLIYLSAMP 416
Db      348 ID-----GFVACTKILGLDELITGLVNDPMHKI PALDPELVNFRVRELTGYLNTQSWP 403
Qy      417 DSLPLDSVFNQLGVIRGRIILANGAVS-LTQGLGISWLGSLRSLGSLALIHNTHLIC 475
Db      404 PHMNFVFNLTITIGRSLYNNGFSLLIMKVLNVTSLGFRSLKEISAGRIYISANQOLC 463
Qy      476 FVHTVPWDQLFRNPQALLHTA-NRPEDECVGEGLACHOLCARHGCMGRPPQCNVCSQF 534
Db      464 YHSLNMTNTRLLRGSEERLDIKYDRPLGECIAEGKVCPLCSSGGCWGGPQCLISCRNY 523
Qy      535 LRQGECEBQRYLQGLPREYVNAHCLPCHPECPQNGSVTFCGREADQCVCAHAKYMDP 594
Db      524 SRGCVCTHGNFIQGEPRFVHQAQCSCHPECLPEGISTYNGSGSDACARCAHFRDGP 583
Qy      595 FCVACRPSGVKRPDLSYMPIMKRPDEBACQPCPINCTHSQVLDLXGCPAEORASPLTS- 653
Db      584 HCVNSCPHILG-AKGPITYKPDVQNECRPCHENCTQCKGPELQDCLGQ-AEVLMSK 639
Qy      654 --IFNNTVSPFLRVKVSASHLEKRRQOKIRKVTMRLLQETLVEPLTSGAMPQAO 711
Db      640 PHVIAVAVGLAVLMLLGSLFYMRRRQNKRAMRYLERGESIEPLDS-EKANKVL 698
Qy      712 MRILKETELRKVKVLGSGAFGTVYKGIWIPGENVKIPVAIKVLRENTSFRANKELIDEA 771
Db      699 ARIFKETELRKVKVLGSGVGFVHKGVWIPGESIKIPVCIKVIDKSGRQSFQAVTDHM 758
Qy      772 YVNAVGSPVYSRLIGLITSTVOLVTOLMPYGLDHYRENRGRISGODLLNMCQIAK 831

```

Db 759 LANGSDHAHIVLLGLCPSSSLQVLTQVLPGLSLDHWKQRETLGPOLLNMGVQIAK 818  
 QY 832 GMSLEEDVRLVHRDLAARNLVKSPNHVKITDFGLARLIDIDETEVHADGKVPKIMAL 891  
 Db 819 GMYYLEESHVVRHDLRLNMYLKSPOGVADFGVADLLPPDDKOLLHBAKPIKIMML 878  
 QY 892 ESILRRRFTHQSDVWSVGYTVWELMTFGAKPYDGIPIARLIPDLLEKGERLPDPPICTIDV 951  
 Db 879 ESIFHGKTYTHQSDVWSVGYTVWELMTFGAEPYAGRLAELPDLEKGERLAQPOLCTIDV 938  
 QY 952 YIMVVKCMIDSECRPFRELVEFSRMARDPQRFVVIQONEDGPASPIDSTFYRSLBD 1011  
 Db 939 YIMVVKCMIDENIRPTEKLANEFTRMARDPRLVIVIKRAS-GPSTP-PAARESVLT 995  
 QY 1012 DDMGDLVDAAEYVLPQGFPCPDPAAGGMVHHRRSSSTSGGDLTLGLEPSEE- 1068  
 Db 996 KEL-GEAELEFEL- DLIDLLEAELEGIA 1021  
 QY 1069 -EAPRSPLAPSEG-AGSDVFDGDLGMGAAGKGLSLPT 1103  
 Db 1022 TSIGSALSLPTGTLTPRSGQLSLSPSSGMPMNQSLDBACLDASVLCGRQFSRPSL 1081  
 QY 1104 HDSPLOKRYSEDPVLPSETDGV-APL-TCT-SPOPE-YVNOP 1143  
 Db 1082 H-PIPRGR-PASESSEGHVTGSEALQEKVYVCRSRSRSPRRDGSAYHSQR 1133  
 QY 1144 DVNPQPPSPREG-LEAPAPAGTILERAATLSP-GKNGV- 1185  
 Db 1134 HSLITVPTPLSPGLLEEDNGVMDTHLRGASRSRBOGLSSVGLSGTEEBED- 1191  
 QY 1186 AFGAVENPEYLTPOGGAAPQPH 1210  
 Db 1192 -EYEMNKRKRGSP-PRP 1209  
 RESULT 11  
 TVEVLV  
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus  
 N:contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase  
 C:Species: avian leukosis virus, ALV  
 C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 11-Jun-1999  
 C/Accession: B00643; A00643  
 C/NILSEN, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Citterenden, L.B.; Raines, M.  
 Cell 41, 719-726, 1985  
 A>Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro  
 A/Reference number: A00643; MUID:85228222; PMID:2988784  
 A/Accession: B00643  
 A/Molecule type: mRNA  
 A/Residues: 1-658 <NIL>  
 A/Cross-references: GB:M10066; GB:M13881; NID:G211749; PIDN:AAA48763.1; PID:G211750  
 A/Note: in Genbank entry CHKEB8B, release 109.0, the source is designated as Gallus gal  
 C/Comment: This protein is synthesized as a gag-env-erbB protein.  
 C/Genetics:  
 A/Gene: gag-env-erbB  
 C/Superfamily: epidermal growth factor receptor; protein kinase homology  
 C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P  
 F:1-6/Product: gag protein (fragment) #status predicted <GAG>  
 F:1-59/Product: env protein (fragment) #status predicted <ENV>  
 F:60-698/Product: protein-tyrosine kinase erbB #status predicted <ERB>  
 F:194-459/Domain: protein kinase homology <KIN>  
 F:202-210/Region: protein kinase ATP-binding motif  
 F:229/Active site: Lys #status predicted  
 Query Match 25.2%; Score 1721.5; DB 1; Length 698;  
 Best Local Similarity 51.5%; Pred. No. 5,6e-64;  
 Matches 368; Conservative 73; Mismatches 151; Indels 123; Gaps 16;  
 QY 578 GEADQCVACAHYKPPFVACPSGVKPDLSYMIWKPPEDEGACQCPINCHSYVDL 637  
 Db 60 GP-DHCKCAHFDIGPHCVKACPAVGENDTL-VWXYADANAVCQCHPCTGCKGCP 116  
 QY 638 DDKGCPAEGRASPLTSIFNNFTVSFWLRLVPKVASHLEKRRQOKRKTMTMRLLQETELV 697

Db 117 GLEGP--NGSKTPSIAGVVGGLCLVWGLIGLYLRRRHIVRKETLRLLQERELV 173  
 QY 698 EPLTPSGAMPQAOGRILKETELRKVYKLGSGAGFTVYKGIWIPGKENVKI-PVAIKVRE 757  
 Db 174 EPLTPSGAMPQAOGRILKETELRKVYKLGSGAGFTVYKGIWIPGKENVKI-PVAIKVRE 233  
 QY 758 NTSPPRANKEIDEAVVMAVGSPPVSRLLGLCLTSVQVLTQVLTQVLPYGLDHYENRGL 817  
 Db 234 ATPSRANKEIDEAVVMAVGSPPVSRLLGLCLTSVQVLTQVLTQVLPYGLDHYENRGL 293  
 QY 818 GSODLMMCMQIAKMSVLEDRVLRDLAARNLVKSPNHVKITDFGLARLIDIDETEV 877  
 Db 294 GSQYLLNMCVOIAGMNYLERRVLRDLAARNLVKSPNHVKITDFGLARLIDIDETEV 353  
 QY 878 HADGKVPKIMMALESILRRRFTHQSDVWSVGYTVWELMTFGAKPYDGIPIAREIPDLLEK 937  
 Db 354 HADGKVPKIMMALESILRRRFTHQSDVWSVGYTVWELMTFGAKPYDGIPIAREIPDLLEK 413  
 QY 938 GERLPQPICTIDVYIMVVKCMIDSECRPFRELVEFSRMARDPQRFVVIQ-NEDIGP 996  
 Db 414 GERLPQPICTIDVYIMVVKCMIDSECRPFRELVEFSRMARDPQRFVVIQ-NEDIGP 473  
 QY 997 ASPLDSTFYRSLDDEDDGDLVDAAEYVLPQGFPCPDPAAGGMVHHRRSSSTSGG 1056  
 Db 474 PSPDTSKRYRLMEEDMEDIVDADELVPHQGF-NSPST- 513  
 QY 1057 GDLTLGLEPSEEA PRSP-APSEGASDVFDGDLGMGAAGKGLSLPTPHDPSPLOR 1111  
 Db 514 -SRTLLSLSATSNNSANTCID-RNCGHVRREDSPQR 551  
 QY 1112 YSEDPVLPSET-DGVVAPLTCSPQPEYVNOGVDRPQPPSPREGPLPAAPAGATLER 1169  
 Db 552 YSSDPTGNFLEESIDDFGL-PAPEVNO-LMPKPS- 585  
 QY 1170 AKTLPQKNGVYKXDF-AFGAVENPEYLTPOGGAAPQPHPAFS 1214  
 Db 586 -TAMQNOQIYNNISLTAISKLPMDSRVQNSHTAINDPEYL-NTQSPFLAK 634  
 QY 1215 PAFNLYYWDQ-DPE- -RGAPPSFTFKGTPTAENEYGLDVP 1254  
 Db 635 TVFESSPYWISQNHQINLDNPDYQDFLPETKPNGLKVPALNEPEYLRVAP 689  
 RESULT 12  
 TYUHL  
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain )  
 C:Species: avian erythroblastosis virus  
 C>Date: 18-Apr-1984 #sequence\_revision 18-Apr-1984 #text\_change 11-Jun-1999  
 C/Accession: A00644; A38022  
 C/Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.  
 Cell 35, 71-76, 1983  
 A>Title: The erbB gene of avian erythroblastosis virus is a member of the src gene  
 A/Reference number: A00644; MUID:84026539; PMID:6313229  
 A/Accession: A00644  
 A/Molecule type: DNA  
 A/Residues: 1-604 <YAM>  
 A/Cross-references: GB:K01246; NID:G209676; PIDN:AAA4400.1; PID:G209678  
 R:Deunire, B.; Henry, C.; Benalissa, M.; Bisette, G.; Clavertie, J.M.; Saule, S.; Mart  
 A/Title: Sequencing the erbA gene of avian erythroblastosis virus reveals a new type  
 A/Reference number: A38022; MUID:84223957; PMID:6328658  
 A/Accession: A38022  
 A/Molecule type: DNA  
 A/Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>  
 A/Cross-references: GB:K02006  
 C/Genetics:  
 A/Gene: erbB  
 C/Superfamily: epidermal growth factor receptor; protein kinase homology  
 C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specif  
 F:130-135/Domain: protein kinase homology <KIN>  
 F:138-146/Region: protein kinase ATP-binding motif  
 F:165/Active site: Lys #status predicted



Query Match 24.3%; Score 1658; DB 1; Length 604;  
 Best Local Similarity 51.4%; Pred. No. 2e-61;  
 Matches 354; Conservative 69; Mismatches 142; Indels 124; Gaps 14;  
 A:Residue: 'A', 832-866 'V', 868-943, 'ONPSLVK' <WAD>  
 A:Cross-references: EMBL:X02293; NID:97922; PIDD:CAA26157.1; PIDD:9929565  
 C/Comment: This sequence is tentative because the introns have not been identified.  
 C/Genetics:  
 A:Gene: FlyBase:Egfr  
 A:Cross-references: FlyBase:FBgn0003731  
 A:Map position: 2.57P  
 C/Superfamily: epidermal growth factor receptor; protein kinase homology  
 C/Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; ph  
 F:1-732/Domain: extracellular #status predicted <EX>  
 F:733-764/Domain: transmembrane #status predicted <TM>  
 F:765-1330/Domain: intracellular #status predicted <INT>  
 F:808-1072/Domain: protein kinase homology <KIN>  
 F:816-824/Region: protein kinase ATP-binding motif  
 F:122-300,324,353,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #stat  
 F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predic  
 F:843/Active site: Lys #status predicted  
 F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status pr

QY 587 CAHYKDPFCVAPCPSPKPDLSYMPKFPDEBGAQCPICNTHSCVDLDDKGPAPQ 646  
 DB 3 CAHIDGPHCVKACAPAGVAGENDTL-VRKYADANAVALCHPNCRCCKGPGEGCP--- 58  
 QY 647 RASPLTIFNNFTVSFRLRPKYSASHLEKRRCKIRKYMRELLOTEVEPLTSGAM 706  
 DB 59 NGSTPSIAAGVGGILCLVAVGIGLYLRRRIYKRLRLLORELEVEPLTSGEA 118  
 QY 707 PNCQMRILKETELRKVKVLSGAFGVYKGIWPDGENYKIPVAIKVLENTSPANKK 766  
 DB 119 PNCNHLKILKETEKVKVLSGAFGVYKGIWPDGENYKIPVAIKVLENTSPANKK 178  
 QY 767 ILDEAYVAGVGSFVSRLLIGICITSTVQLTQMPKGLDHYREKRGSLGQDILNMC 826  
 DB 179 ILDEAYVAGVGSFVSRLLIGICITSTVQLTQMPKGLDHYREKRGSLGQDILNMC 238  
 QY 827 MQAKGMSYLEDVRLVARDLAARVAVKSPNHYKITDPLGLARLDIDETEYHADGKVP 886  
 DB 239 VOIAKGMNLYEERLVARDLAARVAVKSPNHYKITDPLGLARLDIDETEYHADGKVP 1298  
 QY 887 KMALESILRRRTTHQSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLKGERLPQPI 946  
 DB 299 KMALESILRRRTTHQSDWSYGVTVWELMTFGSKPYDGIIPAREIPDLKGERLPQPI 358  
 QY 947 CTIDVYMTKCMWIDECPRPRLVSEFSRMRDQRFVIO-NEEDGAPAPLSTEFY 1005  
 DB 359 CTIDVYMTKCMWIDECPRPRLVSEFSRMRDQRFVIO-NEEDGAPAPLSTEFY 418  
 QY 1006 RSLLEDDMDGLVDAEELVYPOQGFPCPDPAFGAGVHHRHSSSTRSGQDILGLEP 1065  
 DB 419 RSLLEDDMDGLVDAEELVYPOQGFPCPDPAFGAGVHHRHSSSTRSGQDILGLEP 449  
 QY 1066 SEERAPRSPL-----ASESGAGSVDFPDGLGMAKLOSLPTHDEPPLQRYSEDPVPL 1120  
 DB 450 -----SRPLSSLSATSNNSATNCID-----RNGQGHVAREDSFVGRYSDDPGNF 496  
 QY 1121 PSEF--DGYVAPLTCSPQPEYVNPQDVPQPPPREPRLPAPRAGATLBRATLSPGN 1178  
 DB 497 LEESIDGFL-----PAPEVYNO--LMPKKPSTAM----- 524  
 QY 1179 GVYKDVFAF-----GGAVENTEYLTPQGAAPDPPPPAPAFDN 1219  
 DB 525 -VQMOIYVFIISLTAISKLPMDSRVQNSHTAVDNPEYL-----NTNOSPAAKTVFES 575  
 QY 1220 LYYWDODPPERGAPSTFKGPTAENPEY 1248  
 DB 576 SPYWIQSGNHQ-----INIDNPEY 594  
 RESULT 13  
 GQFFE  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbb  
 C:Species: Drosophila melanogaster  
 C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 11-Jun-1999  
 C:Accession: A00640; A38021  
 R:LiVneh, B.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B. Z.  
 Cell 40, 599-607, 1985  
 A:Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding  
 A:Reference number: A00640; MUID:85124611; PMID:2982499  
 A:Accession: A00640  
 A:Molecule type: DNA  
 A:Residues: 1-1330 <Lit>  
 A:Cross-references: EMBL:X03054  
 R:Madsworth, S.C.; Vincent III, W.S.; Bliedean-Wentworth, D.  
 Nature 314, 178-180, 1985  
 A:Title: A Drosophila genomic sequence with homology to human epidermal growth factor re  
 A:Reference number: A38021; MUID:85137938; PMID:2983232  
 A:Accession: A38021

A:Molecule type: DNA  
 A:Residues: 'A', 832-866 'V', 868-943, 'ONPSLVK' <WAD>  
 A:Cross-references: EMBL:X02293; NID:97922; PIDD:CAA26157.1; PIDD:9929565  
 C/Comment: This sequence is tentative because the introns have not been identified.  
 C/Genetics:  
 A:Gene: FlyBase:Egfr  
 A:Cross-references: FlyBase:FBgn0003731  
 A:Map position: 2.57P  
 C/Superfamily: epidermal growth factor receptor; protein kinase homology  
 C/Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; ph  
 F:1-732/Domain: extracellular #status predicted <EX>  
 F:733-764/Domain: transmembrane #status predicted <TM>  
 F:765-1330/Domain: intracellular #status predicted <INT>  
 F:808-1072/Domain: protein kinase homology <KIN>  
 F:816-824/Region: protein kinase ATP-binding motif  
 F:122-300,324,353,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #stat  
 F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predic  
 F:843/Active site: Lys #status predicted  
 F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status pr

QY 80 VQGYVLAHNOYVQPLQRLIRVSTOLF-----EDNALAVLDNGDPLNTPVATGAP 134  
 DB 38 INYVITGLDIPCLTSLYRLQIRGRITFLSVEEEKYALFV-----TY 81  
 QY 135 GGLRELQSLREILKGVLTQORPOLCYQDITLKKDIFHNKQALTLIDNRSRACHP 194  
 DB 82 SKMYTLEIPDLVDVNGVGFHNNVNLGMRITQSEVSNCTDAYVYDFAPRBECK 141  
 QY 195 CSPMKGSRCSWSESESDCSLTITVACAGCA--RCKGPLTDCHEQCAAGCTGPKHSDC 252  
 DB 142 CHESTCHG-CWGEGRKNQCKFSKLTCSPQACAGRCVGRPRECHLFCAGCGTGTQKC 200  
 QY 253 LACLHFNHSGICELHCPALVYNTDTFESMPNPEERYTFGASCCTAACPPNYLSTVGSCT 312  
 DB 201 IACKNFFDEAVSKCECPKPKRYNPTTYLETPEBKRYAVGACVCECP-GHLLRNGACV 259  
 QY 313 LVCPHNOEVTBEDGTORCEKSKPCARVCGVGLGMYIKANSKFIGITEL-----EPAG 366  
 DB 260 RSCPPQDKMDKGE-----CVPNGGCPKTC-----PGVYLAHGNIDSFN 300  
 QY 367 CKKIFGSLAFLESFDG--DPAASNTA-----PIQPEQLQVEFLEITGYLYISAMPDS 418  
 DB 301 CTVIDGNIRILDQTFSGFQDYVANYTWGPRYIPLPERREVESTVKEITGYLYINIEGTHPQ 360  
 QY 419 LPDLSEFQNLQVIRIRILHNGAY-STLQGLQISWLGRLSELSGLALIHNNHLCV 477  
 DB 361 FPNLSFFRLLETTHGQLMESWPAALATVKSLSYSLERKNLKQISSGVVIOHNDLCTV 420  
 QY 478 HTVPMDQLFRNPQALLHTANRBEDEC----- 504  
 DB 421 SNIRWPAIQKEPEQKQWVVENLRADLCGLTLLISVGNIMH:FAICREKMHLLGSV 480  
 QY 505 ----- 504  
 DB 481 QGRLLGSHGVSVPYLQELQFQMLHRLMYLQVINSITDKNSEHQLTACYSPSVPT 540  
 QY 505 -----VG 506  
 DB 541 SUTIRARVAISAGLAMELEITARSASMRSKTLPAERGRVPPWFLGVCASARAGIA 600  
 QY 507 EGLA-----CHQLCARGHCKGPGPTQVNCVSQFLRGQECVEECVLTQGLPREVY---N 556  
 DB 601 EPLAGRAVCRKCHPCELTCTNYGYHCVSKCTHYKREBCCTEC-----PADYITDBE 654  
 QY 557 ARHCLPCHPECCPQNGSVTCRGPADQVCAAHYK-----DPRP-----CVARCSG 603  
 DB 655 QRECTQRPDEC---NG---CTGPADDCGSCNFFLPANETGPVYNSMTENCYSKPLE 708  
 QY 604 VK-PDLSTYMPKFPDEBGAQCPICNTHSCVDLDDKGPAPRASEPLTIFNNFTVSF 662



```

Db      709 MRHVNQYTAIGY-----CAASPRSSKITANLD-----VMMIFITG 747
Qy      663 WLEVPKVASHLEK--RROOKIRKTY--MRLLQETELVEPLTPSGAMPNOQMILKE 717
Db      748 AVLVPTICLCVTVYICRQKQAKKETVKTMAISGREDESELPSPNIGANLCKLIVVD 807
Qy      718 TELRKVKVUGSGAFGVYKGIWIPDGBENKI PVAIKVLRENTSFKANKELIDBAYMAGV 777
Db      808 AELRKGGVLMGMAFGVYKGVWVEGENVKI PVAIKELLKSTGAESESEELREAYIMASE 867
Qy      778 GSPVSRILGICLTSTVOLTOLMPYGCILDHYENRGRGLSGODLLNMCKQIAKGSYLE 837
Db      868 EHVNLKLVAVCSQMLITQLMPGLCLDYVKNRDKISGALLNMSTQLAKGSYLE 927
Qy      838 DVALVHRDLAARNLVK---SPNHVKITDGLARLLIDETEHADGKVPKIMMALES1 894
Db      928 EKLVHRDLAARNVLVRLLAGEDH---DFGLAKLLSSDSNEVKAKGMKPIKMLAECI 983
Qy      895 LRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPORPCTTIVMI 954
Db      984 RNEVFTSKSDVMAFGVYKGIWELTFGQRPHEINPAKDI PDLIEVGLLEQPEICSDIYCT 1043
Qy      955 MVKMMIDSECRPRFRELVEFSRMARDPQRFVVIQNEDLG--PASPLDSTFYRSLLED 1012
Db      1044 LLSGCHLDAMRRTFFQLTVFAFARDPGRYALIGDXTFLPA-----YTSODEKD 1096
Qy      1013 ---DWGDLVDAEELVPQGFPCPPAPGAGVHHRRSSSTRSGGDLTLGLEPSEEE 1069
Db      1097 LIRKLAFTTDSGAIAKPPDYLOPKALGPS-----HRTDCT-----DE 1135
Qy      1070 AP-----HSPLASPAGAGSDVPDG---DLGMAKAGLQSLPTHDSPLQRYSEDPVVL 1120
Db      1136 MPKLNRYCKDPKSNKNSGTGDDERDSSAREVGVNLR-----LDL 1174
Qy      1121 PSETDYVAFLTCSPOPEYVNOQDVPAPPPSPREGPLPAPAPAGATLERAKTLPKNGV 1180
Db      1175 PVDEDDYLP--TQCPGNNNNNNK-----NPNQNNMAVAGVAGYM----- 1214
Qy      1181 VKDVFAPFGAVENPEYL---TPQGAAAPQH-----PPAPSP-AP 1217
Db      1215 --DLIGPVSVNDPEYLLNAQTGVGESPIPTOTIGIPVMGPGTMEVKVPMPSSEPTSS 1272
Qy      1218 DNLVYWD 1224
Db      1273 DHEVYND 1279

```

## RESULT 14

S35745  
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus  
 C:Species: avian erythroblastosis virus  
 C:Date: 03-Mar-1994 #sequence\_revision 26-May-1995 #text\_change 28-Feb-1997  
 C:Accession: S35745  
 R:Vennistrom, B.  
 submitted to the EMBL Data Library, March 1993  
 A:Reference number: S35743  
 A:Accession: S35745  
 A:Molecule type: DNA  
 A:Residues: 1-544 <VEN>  
 A:Cross-references: EMBL:X12707  
 C:Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific F  
 F:135-400/Domain: protein kinase homology <KIN>  
 F:143-151/Region: protein kinase ATP-binding motif  
 F:170/Active site: Lys #stature predicted

Query Match 23.4%; Score 1602; DB 2; Length 544;  
 Best Local Similarity 54.1%; Pred. No. 3,6e-59;  
 Matches 339; Conservative 63; Mismatches 135; Indels 90; Gaps 13;

```

Qy      578 GPEADQVACAHYKDPPECVACSGEVKPDLSYMIWKFPEDEGACQPCPINCTHSCVDL 637
Db      1 GP--DHCKCAHFIDGPHCVKACAPAGVLGENDTL-VKAYADANAACQLCHPNCCTGCKGP 57
Qy      638 DDKGPACQKASPLSTINNNFTGSIWLRVPKVASHLEKQOKIRKTYMRLLQETELV 697
Db      58 GLEGP---NOSKTPSIAAGVVGGLLVVGLGIGLVRHHIRVKRKLRLLOERELV 114
Qy      698 EPLTPSGAMPNOQAKRILKETELRPKVYKLGSGAFGVYKGIWIPDGBENKI PVAIKVLR 757
Db      115 EPLTPSGAMPNOAHRIKLETETFKVKVYKLGSGAFGVYKGIWIPDGBENKI PVAIKVLR 174
Qy      758 NTPSPANKELIDEAVYMAVGSPPVYRLLGICLTSTVOLTOLMPYGCILDHYENRGRGL 817
Db      175 ATSPFANKELIDEAVYMAVSVNPHVCRLGICLTSTVOLTOLMPYGCILDHYENRGRGL 234
Qy      818 GSODLLNMCKQIAKGSYLEVRLVHRLAARNLVKSPNHVKITDGLARLLIDETEHADGK 877
Db      235 GSQVLLNMCKQIAKGMNLYEERHVRDLAARNVLVKTPOHVKITDGLARLLIDETEHADG 294
Qy      878 HADGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEK 937
Db      295 HADGKVPKIMMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEK 354
Qy      938 GERLPORPCTTIVMIWYKCMIDSECRPRFRELVEFSRMARDPQRFVVIQ-NEDIGP 996
Db      355 GERLPORPCTTIVMIWYKCMIDSECRPRFRELVEFSRMARDPQRFVVIQ-NEDIGP 414
Qy      997 ASPLDSTFYRSLLEDMDGDLVDAEELVPQGFPCPPAPGAGVHHRRSSSTRSGG 1056
Db      415 PSFTDSKRYRLMEEDMEDIVDADEVLPQGF-----NSPST----- 454
Qy      1057 GDLTLGLEPSEEARSPPL-----APSEGAGSDVPDGDLGMAKAGLQSLPTHDSPLQR 1111
Db      455 -----SRTLLSLSATSNKSNATNCIDRNG-----H----- 481
Qy      1112 YSEDPVLPBSETDYVAFLTCSPOPEYVNOQDVPAPPPSPREGPLPAPAPAGAT-LEPA 1170
Db      482 -----PYREDFL-----PAPEYVNO--LMPKPSSTAMVNOQYNYISLTAISKL 524
Qy      1171 KTLSPKNGVYKDVAFAGAVENPEYL 1197
Db      525 PIDSRVQY-----SHSTAVNDPEYL 544

```

## RESULT 15

S00727  
 kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi  
 C:Species: avian erythroblastosis virus  
 C:Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 28-Feb-1997  
 C:Accession: S00727  
 R:Scotting, P.; Vennistrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.  
 Oncogene Res. 1, 265-278, 1987  
 A:Title: Common site of mutation in the erbB gene of avian erythroblastosis virus m  
 A:Reference number: S00727; MUID:88217326; PMID:2897102  
 A:Accession: S00727  
 A:Molecule type: DNA  
 A:Residues: 1-545 <SCO>  
 A:Cross-references: EMBL:X06943  
 C:Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; phosphotransferase  
 F:135-400/Domain: protein kinase homology <KIN>  
 F:143-151/Region: protein kinase ATP-binding motif

Query Match 23.3%; Score 1595; DB 2; Length 545;  
 Best Local Similarity 54.1%; Pred. No. 7e-59;  
 Matches 339; Conservative 62; Mismatches 136; Indels 90; Gaps 13;  
 Qy 578 GPEADQVACAHYKDPPECVACSGGVKPDLSYMIWKFPEDEGACQPCPINCTHSCVDL 637  
 Db 1 GP--DHCKCAHFIDGPHCVKACAPAGVLGENDTL-VKAYADANAACQLCHPNCCTGCKGP 57

QY	638	DDXCKPABQRASPLTISFNNTFVSFWLRVPRKVASHLEKREQOKRKXTMRLLOETELV	697
Db	58	GLEBCP---NCSKTPSIIAGVVGGLCLVYVVGILGYLRRRHVTRKTTLRLOERELV	114
QY	698	EPLTPSGAMPNQAOHRIKETELRKYVLGSGAGFTVYKGIWIDGENVKI.PVAIKYLR	757
Db	115	EPLTPSGAPRQAHRIKETELRKYVLGSGAGFTVYKGIWIDGENVKI.PVAIKELRE	174
QY	758	NTSPKANKIIDEAYVMAVGSFVYSRLIGICTSTVOLVTQLMRYGCLLDHVENRGL	817
Db	175	ATSPKANKIIDEAYVMAVSNPHVCRLIGICTSTVOLITQLMRYGCLLDYIREHKONI	234
QY	818	GSODLLNMCQIAGKMSYLEDVRLVHRPLAARNLYVSPNNVKITDPECLALDIDETEX	877
Db	235	GSQYILNMCVOIAGKMYLBERHNLVHRPLAARNLVKIPDVCKITDPECLAQGLADEXEY	294
QY	878	HADGGKVPKIMMALESILRRRFTHQSDWSYGVYWMELMTPGAKPYDIPAREI.PDLLEK	937
Db	295	HAEGGKVPKIMMALESILHRIYTHQSDWSYGVYWMELMTPGSKPYDIPASEISVYLEK	354
QY	938	GERLPQPICTIDYMTIVKCMWIDSECRPPRFRELVSEFSMMARDPQRFVYIQ.-NEDIGP	996
Db	355	GERLPQPICTIDYMTIVKCMWSQADSRPFRELIASFQWADRPRLYLIQOGERHML	414
QY	997	ASPLDSTFYRSLLEDDOMGLVDAAEYLVPOQGFRCPRAPGAGGMMHRRSSSTRAGG	1056
Db	415	PSPTDSKRYRLMEEDMEDYVDAEYLVPHQGF-----NSPST----	454
QY	1057	GDLTLGLEPSEEEAPRSP-----APSEGASDVFDDGLGMAKGLQSLPETHPSPLQF	1111
Db	455	-----SRTPLLSLSLSTSNKSATNCCIDRNG-----H-----	481
QY	1112	YSEDPYVPLPSETQGYVAPLTCSPQPEYVNOVDVROPSPREGFLPARAPAGAT.-LERA	1170
Db	482	-----PVREDDGFL-----PAPEYVNO--LMPKPSSTAWYNOQIYNYISITAIKTL	524
QY	1171	KTLSFGKNGVXKDVFAFGGAVENPEYL.1197	
Db	525	PMDSRYQN-----SHSTAVDNPEYL.544	

Search completed: July 22, 2003, 09:27:28  
Job time : 30.2855 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.642 Seconds

(without alignments)  
4891.279 Million cell updates/sec

Title: SEQ4-632-652-14

Sequence: 1 MELALCRMGLJLALPPCA.....TFKCTPTAENPEYLGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6623	96.9	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5865	85.8	1257	1 ERB2_RAT	P06494 rattus norv
3	5841.5	85.5	1254	1 ERB2_MESAU	P06053 mesocricetu
4	3111	45.5	1210	1 EGFR_HUMAN	P00533 homo sapien
5	3081	45.1	1210	1 EGFR_MOUSE	P00179 mus musculu
6	2920.5	42.7	1308	1 ERB4_HUMAN	Q15303 homo sapien
7	2906	42.5	1308	1 ERB4_MOUSE	Q62856 rattus norv
8	2643.5	38.7	1167	1 XMRK_XIPMA	P13388 xiphiophoru
9	2398.5	35.1	1342	1 ERB3_HUMAN	P21660 homo sapien
10	2325.5	34.0	1339	1 ERB3_RAT	P062799 rattus norv
11	1936	28.3	1426	1 EGFR_DROME	P04412 drosophila
12	1704.5	24.9	634	1 ERB3_ALV	P00534 avian leuko
13	1568	23.3	604	1 ERBB_AVIER	P00535 avian eryth
14	1585	22.7	540	1 ERBB_AVIEU	P11373 avian eryth
15	1550	18.6	1333	1 EGFR_CHICK	P13387 gallus gall
16	1274	16.7	245	1 ERB2_MOUSE	P24348 caenorhabdi
17	1142.5	10.4	1363	1 ILPR_BRALA	P70324 mus musculu
18	713	10.1	1372	1 INSR_MOUSE	P02466 brachiolesto
19	693.5	10.1	1302	1 INSR_MOUSE	P15208 mus musculu
20	692	10.1	1302	1 INSR_MOUSE	Q94514 mus musculu
21	691	10.1	1302	1 INSR_MOUSE	Q94514 mus musculu
22	680	10.1	1383	1 INSR_HUMAN	P05127 rattus norv
23	687	10.1	1297	1 IRR_HUMAN	P14616 homo sapien
24	680.5	10.0	1300	1 IRR_HUMAN	P14617 cavia porce
25	680	10.0	1477	1 HTRK_HYDAT	Q25197 hydra atten
26	680	10.0	1607	1 MIRP_LYST	P08069 homo sapien
27	644	9.4	1367	1 IGR_HUMAN	P08069 homo sapien
28	630	9.2	1373	1 IGR_MOUSE	P24062 rattus norv
29	626.5	9.2	1370	1 IGR_RAT	Q60751 mus musculu
30	617	9.0	1390	1 INSR_AEDAE	Q93105 aedes aegypt
31	616	9.0	1246	1 INSR_DROME	P09208 drosophila
32	591	8.6	1114	1 RET_HUMAN	P07949 homo sapien
33	589.5	8.6	1053	1 FAK1_CHICK	Q00344 gallus gall

34	582	8.5	757	1 HT16_HYDAT	P53156 hydra atten
35	579	8.5	1052	1 FAK1_MOUSE	P34152 mus musculu
36	578.5	8.5	984	1 EPB1_CHICK	P07494 gallus gall
37	576.5	8.4	984	1 EPB1_RAT	P09759 rattus norv
38	576	8.4	1055	1 FAK1_RAT	Q03346 rattus norv
39	572	8.4	1052	1 FAK1_HUMAN	Q05397 homo sapien
40	570.5	8.3	984	1 EPB1_HUMAN	P54762 homo sapien
41	570	8.3	987	1 EPB4_HUMAN	P54760 homo sapien
42	569	8.3	902	1 EPB8_XENLA	Q91736 xenopus lae
43	569	8.3	1068	1 FAK1_XENLA	Q91738 xenopus lae
44	555.5	8.1	1009	1 FAK2_RAT	P70600 rattus norv
45	554.5	8.1	746	1 ABL_MLVAB	P00521 abelson mur

## ALIGNMENTS

RESULT 1  
ID ERB2\_HUMAN STANDARD: PRT; 1255 AA.  
AC P04626;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
DE (P13387B2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell  
GN ERB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid=9606;  
RN (1)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86118663; PubMed=3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,  
RA Saito T., Toyoshima K.,  
RT "Similarity of protein encoded by the human c-erb-B-2 gene to  
RT epidermal growth factor receptor.";  
RL Nature 319:230-234(1986).  
RN (2)  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86070181; PubMed=2999974;  
RA Cousens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,  
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,  
RA Francke U., Levinson A., Ullrich A.,  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor  
RT shares chromosomal location with neu oncogene.";  
RL Science 230:1132-1139(1985).  
RN (3)  
RP SEQUENCE OF 737-1031 FROM N.A.  
RX MEDLINE=86016729; PubMed=295967;  
RA Semba K., Kamata N., Toyoshima K., Yamamoto T.,  
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the  
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a  
RL human salivary gland adenocarcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).  
RN (4)  
RP VARIANTS VAL-654 AND VAL-655.  
RX MEDLINE=93194196; PubMed=8095488;  
RA Ehsani A., Low J., Wallace R.B., Wu A.M.,  
RT "Characterization of a new allele of the human ERB2 gene by allele-  
RT specific competition hybridization.";  
RL Genomics 15:426-429(1993).  
CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
CC ALTHOUGH NEUREGULIN DO NOT INTERACT WITH IT ALONE. GP30 IS A  
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
CC ALPHA AND AMPHIREGULIN.  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate  
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC (POTENTIAL).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY).  
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN  
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M11767; AAA35808.1; -.  
 DR EMBL; M11761; AAA35808.1; JOINED.  
 DR EMBL; M11762; AAA35808.1; JOINED.  
 DR EMBL; M11763; AAA35808.1; JOINED.  
 DR EMBL; M11764; AAA35808.1; JOINED.  
 DR EMBL; M11765; AAA35808.1; JOINED.  
 DR EMBL; M11766; AAA35808.1; JOINED.  
 DR EMBL; M11730; AAA35493.1; -.  
 DR EMBL; M12036; AAA35878.1; -.  
 DR EMBL; X03563; CAA27060.1; -.  
 DR PIR; A25491; A25491.  
 DR PIR; A24571; A24571.  
 DR HSSP; P11362; 1FGK.  
 DR Genem; HGNC:3430; ERBB2.  
 DR MIM; 164870; -.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_Pkinase.  
 DR InterPro; IPR004019; YLP\_motif.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR Pfam; PF02757; YLP; 2.  
 DR ProDom; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SM00261; Tyrc; 3.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Multigene Family; Receptor; Signal;  
 DR Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 DR Polymorphism.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 653 675 POTENTIAL.  
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 720 987 PROTEIN KINASE.  
 FT BINDING 726 734 ATP (BY SIMILARITY).  
 FT ACT\_SITE 753 753 ATP (BY SIMILARITY).  
 FT DISULFID 845 845 BY SIMILARITY.  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 220 227 BY SIMILARITY.  
 FT DISULFID 224 235 BY SIMILARITY.  
 FT DISULFID 236 244 BY SIMILARITY.  
 FT DISULFID 240 252 BY SIMILARITY.  
 FT DISULFID 255 264 BY SIMILARITY.  
 FT DISULFID 268 295 BY SIMILARITY.  
 FT DISULFID 299 311 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 511 520 BY SIMILARITY.  
 FT DISULFID 515 528 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 567 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 124 68 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 124 68 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARIANT 654 654 /FTID=VAR\_004077.  
 FT VARIANT 655 655 I->V.  
 FT CONFLICT 1170 1170 /FTID=VAR\_004078.  
 FT SEQUENCE 1255 AA; 137909 MW; 39E9DFDA04C9F62 CRC64;  
 SQ

Query Match 96.9%; Score 6623; DB 1; Length 1255;  
 Best Local Similarity 97.3%; Pred. No. 0; Mismatches 28; Indels 0; Gaps 0;  
 Matches 1221; Conservative 6;

QY 1 MELIACRMLLALPFGASTQVCTGDMRLPASPETHLDMRLHLYGCGVQGNL 60  
 DB 1 MELIACRMLLALPFGASTQVCTGDMRLPASPETHLDMRLHLYGCGVQGNL 60  
 QY 61 ELTYLPTNASTSFLDIOGVGVLIANQVQVLPRLRYRGVQLFEDNYALVDNG 120  
 DB 61 ELTYLPTNASTSFLDIOGVGVLIANQVQVLPRLRYRGVQLFEDNYALVDNG 120  
 QY 121 DPANNTPTVGTASPGGLRELQRLSTELIKGVLLQRPOLCYQDTILMKDFFHNQOLA 180  
 DB 121 DPANNTPTVGTASPGGLRELQRLSTELIKGVLLQRPOLCYQDTILMKDFFHNQOLA 180  
 QY 121 DPANNTPTVGTASPGGLRELQRLSTELIKGVLLQRPOLCYQDTILMKDFFHNQOLA 180  
 DB 121 DPANNTPTVGTASPGGLRELQRLSTELIKGVLLQRPOLCYQDTILMKDFFHNQOLA 180  
 QY 181 LTLIDNRSRACHPCSPMKCSRGWSESESDCSLTRYVACGACARCKPFLTDCHEQC 240  
 DB 181 LTLIDNRSRACHPCSPMKCSRGWSESESDCSLTRYVACGACARCKPFLTDCHEQC 240  
 QY 181 LTLIDNRSRACHPCSPMKCSRGWSESESDCSLTRYVACGACARCKPFLTDCHEQC 240  
 DB 181 LTLIDNRSRACHPCSPMKCSRGWSESESDCSLTRYVACGACARCKPFLTDCHEQC 240  
 QY 241 AAGCTGPRKSDCLACLFHNSGICGLHCPALVTYNTDFESNPNBEGRTFASCVTACP 300  
 DB 241 AAGCTGPRKSDCLACLFHNSGICGLHCPALVTYNTDFESNPNBEGRTFASCVTACP 300  
 QY 241 AAGCTGPRKSDCLACLFHNSGICGLHCPALVTYNTDFESNPNBEGRTFASCVTACP 300  
 DB 241 AAGCTGPRKSDCLACLFHNSGICGLHCPALVTYNTDFESNPNBEGRTFASCVTACP 300  
 QY 301 VNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFGIT 360  
 DB 301 VNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFGIT 360  
 QY 301 VNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFGIT 360  
 DB 301 VNYLSTDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCARVCYGLGMQYIKANSKFGIT 360  
 QY 361 ELIEFACCKKIFGSLAFPEPSFGDPASNTAPLQEPQLOVFEETLITGVLYISAMPDILP 420  
 DB 361 ELIEFACCKKIFGSLAFPEPSFGDPASNTAPLQEPQLOVFEETLITGVLYISAMPDILP 420  
 QY 361 IOEFACCKKIFGSLAFPEPSFGDPASNTAPLQEPQLOVFEETLITGVLYISAMPDILP 420  
 DB 361 IOEFACCKKIFGSLAFPEPSFGDPASNTAPLQEPQLOVFEETLITGVLYISAMPDILP 420  
 QY 421 DLSVQNLQVIGRIHNGAVSLTQGGISMLGLRSRLREIGSGALALHHNTHLFCFVTV 480  
 DB 421 DLSVQNLQVIGRIHNGAVSLTQGGISMLGLRSRLREIGSGALALHHNTHLFCFVTV 480  
 QY 421 DLSVQNLQVIGRIHNGAVSLTQGGISMLGLRSRLREIGSGALALHHNTHLFCFVTV 480  
 DB 421 DLSVQNLQVIGRIHNGAVSLTQGGISMLGLRSRLREIGSGALALHHNTHLFCFVTV 480  
 QY 481 PWDQLFRRNHQALLHTANRPEDECVGEGIALCHQLCARHGWPGPTQCVNSQFLRGEC 540  
 DB 481 PWDQLFRRNHQALLHTANRPEDECVGEGIALCHQLCARHGWPGPTQCVNSQFLRGEC 540  
 QY 481 PWDQLFRRNHQALLHTANRPEDECVGEGIALCHQLCARHGWPGPTQCVNSQFLRGEC 540  
 DB 481 PWDQLFRRNHQALLHTANRPEDECVGEGIALCHQLCARHGWPGPTQCVNSQFLRGEC 540  
 QY 541 VEECVLQCLPREYVNAHCHLPCHEQCPQNSVTCFPEADQCAACHYDPPCVARC 600  
 DB 541 VEECVLQCLPREYVNAHCHLPCHEQCPQNSVTCFPEADQCAACHYDPPCVARC 600  
 QY 541 VEECVLQCLPREYVNAHCHLPCHEQCPQNSVTCFPEADQCAACHYDPPCVARC 600  
 DB 541 VEECVLQCLPREYVNAHCHLPCHEQCPQNSVTCFPEADQCAACHYDPPCVARC 600  
 QY 601 PEGVAPDLSYMBIMKFPDEGACQPCPINCTHSCVDLDDKCPAQRASPLTSIFNNFTV 660  
 DB 601 PEGVAPDLSYMBIMKFPDEGACQPCPINCTHSCVDLDDKCPAQRASPLTSIFNNFTV 660  
 QY 601 PEGVAPDLSYMBIMKFPDEGACQPCPINCTHSCVDLDDKCPAQRASPLTSIFNNFTV 660  
 DB 601 PEGVAPDLSYMBIMKFPDEGACQPCPINCTHSCVDLDDKCPAQRASPLTSIFNNFTV 660  
 QY 661 SFWLRPKYSASHLEKROQKTRKTYMRLLQETLVEPLTPSGAMPNQAQRIKTEL 720  
 DB 661 ILVVVLGVFGILLKROQKTRKTYMRLLQETLVEPLTPSGAMPNQAQRIKTEL 720

QY	721	RAKVLKLGSAFETVYKGIIPDGENKIVAIKVLRENSPANKIIDEAIVMGVSP	780
Db	721	RKVKVLGSAFETVYKGIIPDGENKIVAIKVLRENSPANKIIDEAIVMGVSP	780
QY	781	YVSRLLGICLTSTVQLVLTOLMPYGCILDHVENRGLSGODLLNCMOIAKMSYLEDR	840
Db	781	YVSRLLGICLTSTVQLVLTOLMPYGCILDHVENRGLSGODLLNCMOIAKMSYLEDR	840
QY	841	LVHRDLAARNVLVKS PNHYKIDFGARLLIDETETHADGKVPDKMALESILRRPT	900
Db	841	LVHRDLAARNVLVKS PNHYKIDFGARLLIDETETHADGKVPDKMALESILRRPT	900
QY	901	HOSDVMSGVTVWELMTSGAKPYDGIPAEIPDLLKGRLEDPQICITIDVYMIWCKMM	960
Db	901	HOSDVMSGVTVWELMTSGAKPYDGIPAEIPDLLKGRLEDPQICITIDVYMIWCKMM	960
QY	961	IDSECRPRERELVSERSRAARDPQRFVVIQNEIDLPAISLSTFTFRSLLEDMDMDLDA	1020
Db	961	IDSECRPRERELVSERSRAARDPQRFVVIQNEIDLPAISLSTFTFRSLLEDMDMDLDA	1020
QY	1021	EEYLVPOQGFPCPDPAFAGAGVWHRRHSSSTRSGGDLITGLBPSSEAPRSLAPSEG	1080
Db	1021	EEYLVPOQGFPCPDPAFAGAGVWHRRHSSSTRSGGDLITGLBPSSEAPRSLAPSEG	1080
QY	1081	AGSDVPDGGILGMAKGLQSLDPTHPSPLQRYSEDPVPLPSETGYVAPLTCSQPEYV	1140
Db	1081	AGSDVPDGGILGMAKGLQSLDPTHPSPLQRYSEDPVPLPSETGYVAPLTCSQPEYV	1140
QY	1141	NQPDVRRPOPSPRESEPLPAAPAGATLEBPKTLISGKXGVVDVPAFGAVENPEYLRQ	1200
Db	1141	NQPDVRRPOPSPRESEPLPAAPAGATLEBPKTLISGKXGVVDVPAFGAVENPEYLRQ	1200
QY	1201	GGAAPQHPHPAPFSAFNLTYWDDPBERGAPSTFEGTPTAENPEYLGIDVPY	1255
Db	1201	GGAAPQHPHPAPFSAFNLTYWDDPBERGAPSTFEGTPTAENPEYLGIDVPY	1255
RESULT 2			
ERR2 RAT			
ID	ERR2 RAT	STANDARD:	PRT: 1257 AA.
AC	P06454:		
DT	01-JAN-1988 (Rel. 06, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)		
DE	(p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).		
GN	ERBB2 OR NEU		
OS	Rattus norvegicus (Rat).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.		
OX	NCBI_TaxID=10116;		
RN	[1]		
RN	SEQUENCE FROM N.A.		
RC	TISSUE=Neuroblastoma;		
RX	MEDLINE=86118662; Pubmed=3945311;		
RA	Bagmann C.I., Hung M.-C., Weinberg R.A.;		
RT	"The neu oncogene encodes an epidermal growth factor receptor-related protein.";		
RL	Nature 319:226-230 (1986).		
RN	[2]		
RN	SEQUENCE OF 852-905 FROM N.A.		
RP	TISSUE=Sciatic nerve;		
RC	MEDLINE=91222560; Pubmed=2025425;		
RA	Lai C., Lemke G.;		
RT	"An extended family of protein-tyrosine kinase genes differentially		
RT	expressed in the vertebrate nervous system.";		
RL	Neuron 6:691-704 (1991).		
RN	[3]		
RP	STRUCTURE BY NMR OF 650-668.		
RX	MEDLINE=92155181; Pubmed=1346763;		
RA	Gullick W.J., Bottomley A.C., Lotfs F.U., Doak D.G., Mulvey D.,		

[illegible]

```

FT DISULFID 569 586 BY SIMILARITY.
FT DISULFID 589 598 BY SIMILARITY.
FT DISULFID 602 625 BY SIMILARITY.
FT DISULFID 628 636 BY SIMILARITY.
FT DISULFID 632 644 BY SIMILARITY.
FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).
SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 85.8%; Score 5665; DB 1; Length 1257;
Best Local Similarity 86.2%; Pred. No. 1,66-302;
Matches 1084; Conservative 48; Mismatches 123; Indels 2; Gaps 2;

QY 1 MELALCRWGLLLALLPGAASTOCTGDMKRLPASPETHLDMRLHYQCQVVGQNL 60
DB 1 MELAAWCRKGFLLALLPGLIAGTQCTGDMKRLPASPETHLDMRLHYQCQVVGQNL 60
QY 61 ELTYPTNASLFLQIDQEVQGVLIANQVQVQLRLRYRGQLPEEDNALAVLDNG 120
DB 61 ELTYVPANASLFLQIDQEVQGVLIANQVQVQLRLRYRGQLPEEDNALAVLDNR 120
QY 121 DELNNTPTVT-GASPGGLREQLRSLTELKGVLIQNRNPOLCYODTLMKDIPKNNQL 179
DB 121 DQDNVAATPRTGREGLELDRLSTELKGVLIQNRNPOLCYODTLMKDIPKNNQL 180
QY 180 ALTLIDTNSRACHPCSPCKGSRGWSSESDCQSLRTVCAGCARCKGLPTDCCHEQ 239
DB 181 AFDVIDTNSRACPAPACKKNHGMWSESDCQSLRTVCAGCARCKGLPTDCCHEQ 240
QY 240 CAAGCTGPHGSDCLACIHNHSGICELHCPALVTNTPTFSEMPREBYTFGASCYATC 299
DB 241 CAAGCTGPHGSDCLACIHNHSGICELHCPALVTNTPTFSEMPREBYTFGASCYATC 300
QY 300 PNYLSTDVSGCTLVCPHNOEVTAEADGTORCEKSKPCARCYGLGNQYIKANSKFTGI 359
DB 301 PNYLSTEVSGCTLVCPHNOEVTAEADGTORCEKSKPCARCYGLGNHMLRGAATISD 360
QY 360 TELERFAGCKIRGSLAFLESPDOPASTAPLOEOLQVFTLEITIGYLIAMPSL 419
DB 361 NVQEPDGGCKIRGSLAFLESPDOPASTAPLOEOLQVFTLEITIGYLIAMPSL 420
QY 420 PDLSTFONLOVIRGRILHNGAYSLTLOGLISMLGRSLRELGSGLALIHNNTHLCFVHT 479
DB 421 RDLSTFONRIIRGRILHNGAYSLTLOGLISMLGRSLRELGSGLALIHNNTHLCFVHT 480
QY 480 VWMDOLEFNFHQAALHTANRPDE-CVGEGLACHOLCARHGWGFGPTQVCNCSQFLRGQ 538
DB 481 VWMDOLEFNFHQAALHTANRPDE-CVGEGLACHOLCARHGWGFGPTQVCNCSQFLRGQ 540
QY 539 ECVESCRVLOGLPREVYVARHCTLPCHPGEPONGSVTGPBPADCCVCAHAKDPPEVVA 598
DB 541 ECVESCRVWKGIPREYVSDKCLPCHPGEPONGSVTGPBPADCCVCAHAKDPPEVVA 600
QY 599 RCPGSGVPLSLYMPIMKPFDEEGACQPCINCTHSCVDLDDKGPABOASPLTSPNNF 658
DB 601 RCPGSGVPLSLYMPIMKPFDEEGACQPCINCTHSCVDLDDKGPABOASPLTSPNNF 660
QY 659 TVSFMLRVKVSASHLEKRRROOKIRKYMRLQSTELVEPLTFSGAMPNOAQMKILET 718
DB 661 VGVLLFLIVVAVGLIKRRROKIRKYMRLQSTELVEPLTFSGAMPNOAQMKILET 720
QY 719 ELRKXKVLGSGAGFVYKGIWIPDENYKIPVAIVLENTSPKANKIILDEAYVMAVG 778
DB 721 ELRKXKVLGSGAGFVYKGIWIPDENYKIPVAIVLENTSPKANKIILDEAYVMAVG 780
QY 779 SFYVSRLLGICLTSTVQLVTOQLMPYGCILLDHVRENRGRIGSODLLNMCQIAKGNSTYED 838

```

```

DB 781 SFYVSRLLGICLTSTVQLVTOQLMPYGCILLDHVRENRGRIGSODLLNMCQIAKGNSTYED 840
QY 839 VRLVHRLARNLVNSPNNVKTITDFCLALLDIDETRYADGKVPKMALESILRR 898
DB 841 VRLVHRLARNLVNSPNNVKTITDFCLALLDIDETRYADGKVPKMALESILRR 900
QY 899 FTHOSDWSGVTWELMTFGAKPYDGPRIEDLEKGERLPQPICTIDVYMIWKC 958
DB 901 FTHOSDWSGVTWELMTFGAKPYDGPRIEDLEKGERLPQPICTIDVYMIWKC 960
QY 959 WMIDSECRPFRELVSFSSMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLV 1018
DB 961 WMIDSECRPFRELVSFSSMARDPQRFVVIQNEIDLGPASPLDSTFYRSLLEDMDGLV 1020
QY 1019 DAERYLVPOQGFPPDPAPAGAMVHHRHSSSTRSGGDLTGLSESEEARSPAPS 1078
DB 1021 DAERYLVPOQGFPPDPAPAGAMVHHRHSSSTRSGGDLTGLSESEEARSPAPS 1080
QY 1079 EGAGSDVFDGDLGKAAGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQE 1138
DB 1081 EGAGSDVFDGDLGKAAGLQSLPTHDPSPLQRYSEDPVPLPSETDGYVAPLTCSPQE 1140
QY 1139 YVNOQDVRPQPPSEPEPLPAAPACATLERATLSFGKGVYKVDVAFGAVENPEYLT 1198
DB 1141 YVNOSEVQPPPLPPEGLPLPVPAGATLERPXTLSGKGVYKVDVAFGAVENPEYLT 1200
QY 1199 POGGAAPPPPPAFSPAFNLYYVODDPERGAPSTFPGTPTAENPEYLGADVPV 1255
DB 1201 PREGTAPPPPPAFSPAFNLYYVODDPERGAPSTFPGTPTAENPEYLGADVPV 1257

RESULT 3
ERB2_MESAU STANDARD; PRT; 1254 AA.
ID ERB2_MESAU
AC 060553;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).
GN ERB2 OR NEU
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_Taxid=10036;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Nerve;
RX MEDLINE=94193007; Pubmed=7908275;
RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
RA Yamazaki Y., Ishikawa T.;
RT "Cloning and activation of the Syrian hamster neu proto-oncogene.";
RL Gene 140:251-255 (1994).
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. G930 IS A
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- PFM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>)  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch).

DR EMBL; D16295; BAA03801.1; -  
 DR HSBP; P11362; IFGK.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR InterPro; IPR004019; YLP\_motif.  
 DR Pfam; PF00069; kinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR Pfam; PF02757; YLP; 2.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00261; Fc; 3.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR Transmembrane; G1ycoprotein; Multigene family; Receptor; Signal;  
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Proto-oncogene; Disease mutation.  
 FT SIGNAL 1 21  
 FT CHAIN 22 1254  
 FT DOMAIN 22 652 RECEPTOR PROTEIN-TYROSINE KINASE EBBB-2.  
 FT TRASMEN 653 675 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 676 1254 POTENTIAL.  
 FT DOMAIN 158 368 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 472 644 CYS-RICH.  
 FT DOMAIN 720 987 PROTEIN KINASE.  
 FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
 FT BINDING 753 753 ATP (BY SIMILARITY).  
 FT ACT\_SITE 845 845 BY SIMILARITY.  
 FT DISULFID 189 204 BY SIMILARITY.  
 FT DISULFID 189 212 BY SIMILARITY.  
 FT DISULFID 236 244 BY SIMILARITY.  
 FT DISULFID 240 252 BY SIMILARITY.  
 FT DISULFID 255 264 BY SIMILARITY.  
 FT DISULFID 268 295 BY SIMILARITY.  
 FT DISULFID 299 311 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 511 520 BY SIMILARITY.  
 FT DISULFID 515 528 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 544 560 BY SIMILARITY.  
 FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 567 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 642 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 125 125 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).  
 FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).  
 SO SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;

Query Match 85.5%; Score 5841.5; DB 1; Length 1254;  
 Best Local Similarity 85.7%; Pred. No. 2.6e-301;  
 Matches 1076; Conservative 57; Mismatches 121; Indels 1; Gaps 1;

QY 1 MELALCRWGLLLALLPPGASTVCTGTDKMLLPASPTHLDMLRHLVGGCVVQGNL 60  
 DB 1 MELAAWCGWGLLLALLSPGASGTCTGTDKMLLPASPTHLDIVHLLVGGCVVQGNL 60

QY 61 ELTYLPTNASTLSPFODIOEVGYVLIANNCVROVPLORLRIVRGTOLEFEDVYALAVLNDG 120  
 DB 61 ELTYLPANATLSFQDIOEVGYVLIANSQVRAHPFLQRLRIVRGTOLEFEDVYALAVLNDR 120  
 QY 121 DPLNNTPTVTASGGLRELOSLTEILKGVLIQRNPOLCYODTLKMDIFKKNQOLA 180  
 DB 121 DPLNNTPTVTASGGLRELOSLTEILKGVLIQRNPOLCYODTLKMDIFKKNQOLA 180  
 QY 181 LTLIDTRSRACHCSPMCKSRGSSSESCOSTLTVCAGGACRKGPLPTCCHEQC 240  
 DB 181 LTLIDTRSRACHCSPMCKSRGSSSESCOSTLTVCAGGACRKGPLPTCCHEQC 240  
 QY 241 AAGCTGPRHSDCLALFHNSGICELHCPALVTYNTDFFSMPNREGRYFGASCYACP 300  
 DB 241 AAGCTGPRHSDCLALFHNSGICELHCPALVTYNTDFFSMPNREGRYFGASCYACP 300  
 QY 301 YNVLSTVGSTLVCPLHNOGYVAEDGTQCEKSKRCARVCYGLNQYIKANSKFTGIT 360  
 DB 301 YNVLSTVGSTLVCPLHNOGYVAEDGTQCEKSKRCARVCYGLNQYIKANSKFTGIT 360  
 QY 361 ELBPAGCKKIFGSLAFIPESFDDGPASNTAPLOPEQLOVEFTEIETGYLYISAMPDLP 420  
 DB 361 ELBPAGCKKIFGSLAFIPESFDDGPASNTAPLOPEQLOVEFTEIETGYLYISAMPDLP 420  
 QY 421 DLSYFQNLQVIRGRIHNGAYSLTLQGLISWIGLSRLBELSGGLAIHNNTHLCFVHTV 480  
 DB 421 DLSYFQNLQVIRGRIHNGAYSLTLQGLISWIGLSRLBELSGGLAIHNNTHLCFVHTV 480  
 QY 481 PMDDLFPNPHOALHTNRPEDCEVSGGLACHOLCARGHWGEGPTQCVNCSQFLRQEC 540  
 DB 481 PMDDLFPNPHOALHTNRPEDCEVSGGLACHOLCARGHWGEGPTQCVNCSQFLRQEC 540  
 QY 541 VEBCKVYQGLPREYVNAHCLPCHPECOPNGSVTCGPADQCVACAHYKDPFCVYAR 600  
 DB 541 VEBCKVYQGLPREYVNAHCLPCHPECOPNGSVTCGPADQCVACAHYKDPFCVYAR 600  
 QY 601 PSQVYKPLSTMPYKPFDEGACQPCINCTHSCVDLDKGCFAEQRASPLTJIFNNFTV 660  
 DB 601 PSQVYKPLSTMPYKPFDEGACQPCINCTHSCVDLDKGCFAEQRASPLTJIFNNFTV 660  
 QY 661 SFMLRVPKVSASHLEKRRQOKIRKYMTRRLQETELVEPLTPSGAMPNQOMKILKETEL 720  
 DB 661 SFMLRVPKVSASHLEKRRQOKIRKYMTRRLQETELVEPLTPSGAMPNQOMKILKETEL 720  
 QY 721 RRYKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLRBENTSPKANKEILDEAYVMAGVSP 780  
 DB 721 RRYKVLGSGAFGVYKGIWIPDGENYKIPVAIKVLRBENTSPKANKEILDEAYVMAGVSP 780  
 QY 781 YVSRLLGICLSTVQVLTQMPYGCILLDHYRENRGRLGSDLLNCKMQIAKMSYLEDVR 840  
 DB 781 YVSRLLGICLSTVQVLTQMPYGCILLDHYRENRGRLGSDLLNCKMQIAKMSYLEDVR 840  
 QY 841 LVHRDLAARNVLVKSNHVKITDFGLARLLIDETEHYADGGKVPKIMMALESLIRRF 900  
 DB 841 LVHRDLAARNVLVKSNHVKITDFGLARLLIDETEHYADGGKVPKIMMALESLIRRF 900  
 QY 901 HQSDVWSYGTWELMTFGAKPYDGIIPAREIPDLLEKGRLLPOPICTIDYIMVYKCM 960  
 DB 901 HQSDVWSYGTWELMTFGAKPYDGIIPAREIPDLLEKGRLLPOPICTIDYIMVYKCM 960  
 QY 961 IDSECRPRPRELVSEFSRMAPDORFVUQNEDLGASPLDSTFYSLLEDMDGLVNA 1020  
 DB 961 IDSECRPRPRELVSEFSRMAPDORFVUQNEDLGASPLDSTFYSLLEDMDGLVNA 1020  
 QY 1021 EBYLVYQGGFCDDPAAGGVMVHHRHSSSTRSGGDLTLGLBSEEBEAPRSLAPSE 1080  
 DB 1021 EBYLVYQGGFCDDPAAGGVMVHHRHSSSTRSGGDLTLGLBSEEBEAPRSLAPSE 1080  
 QY 1081 AGSDVFEGLGKATKGPQISIRPLQRYSEDPVLPPLPETGQYVAPLACSPQPEV 1140  
 DB 1081 AGSDVFEGLGKATKGPQISIRPLQRYSEDPVLPPLPETGQYVAPLACSPQPEV 1140

QY 1141 NOPDVROPSPREGLPLPAPAGATLERAKTSLSPGKGVVQVDFAGAVENPEYLTPO 1200  
 DB 1141 NOPDVROPSPREGLPLPAPAGATLERAKTSLSPGKGVVQVDFAGAVENPEYLTPO 1200  
 QY 1201 GGAAQPPPPAFAFDNLYYWDOPPERGAPSTFKGTPTAENPEYLGIDVPY 1255  
 DB 1201 GGAAQPPPPAFAFDNLYYWDOPPERGAPSTFKGTPTAENPEYLGIDVPY 1255  
 RESULT 4  
 EGFR\_HUMAN STANDARD; PRT; 1210 AA.  
 ID EGFR\_HUMAN  
 AC P00533; P06268; Q14225; Q9UMD8; Q9JMG5; Q92795; Q00732;  
 AC Q00688; Q9B2S2; Q9H2C9; Q9GZX1; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01. Created)  
 DT 01-NOV-1997 (Rel. 35. Last sequence update)  
 DT 15-JUN-2002 (Rel. 41. Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor  
 protein-tyrosine kinase ErbB-1).  
 GN EGFR OR ERBB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN 11  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=84219729; PubMed=6328312;  
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
 RA Lee J., Varden Y., Liberman T.A., Schlessinger J., Downward J.,  
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;  
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant  
 RT expression of the amplified gene in A431 epidermoid carcinoma cells."  
 RT Nature 309:418-425 (1984).  
 RL [2]  
 RN 12  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=95382957; PubMed=7654368;  
 RA Ilekis J.V., Stark B.C., Scoccia B.;  
 RT "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta."  
 RT Mol. Reprod. Dev. 41:149-156 (1995).  
 RL [3]  
 RN 13  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=97078686; PubMed=8918811;  
 RA Reiter U.L., Maibach N.J.;  
 RT "A 1.8 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor."  
 RT Nucleic Acids Res. 24:4050-4056 (1996).  
 RL [4]  
 RN 14  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=97256547; PubMed=9103388;  
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;  
 RT "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGFR) in ovarian cancer."  
 RT Gynecol. Oncol. 65:36-41 (1997).  
 RL [5]  
 RN 15  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RX TISSUE=Placenta;  
 RX MEDLINE=21100872; PubMed=11161793;  
 RA Reiter U.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
 RA Schehl Sinclair C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGFR transcripts encoding truncated receptor  
 RT isoforms."  
 RT Genomics 71:1-20 (2001).  
 RL [6]  
 RN 16  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reiter U.L., Thredgill D.W., Danielsen A.J., Schehl C.M.,  
 RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,  
 RA Maibach N.J.;  
 RT "Human and mouse alternative EGFR transcripts encoding only the  
 RT extracellular domain of the receptor."  
 RT Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.  
 RL [7]  
 RN 17  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,  
 RA Evans R.M., Verna I.M., Gill G.N., Rosenfield M.G.;  
 RT "Expression cloning of human EGF receptor complementary DNA: gene  
 RT amplification and three related messenger RNA products in A431  
 RT cells."  
 RT Science 224:843-848 (1984).  
 RL [8]  
 RN 18  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.;  
 RT "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAs overproduced in A431 carcinoma cells."  
 RT Nature 309:806-810 (1984).  
 RL [9]  
 RN 19  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,  
 RA O'Malley B.W.;  
 RT "Isolation of an evolutionarily conserved epidermal growth factor  
 RT receptor cDNA from human A431 carcinoma cells."  
 RT Biochem. Biophys. Res. Commun. 124:125-132 (1984).  
 RL [10]  
 RN 20  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kinchington D., Ullrich A.,  
 RA Waterfield M.D.;  
 RT "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription."  
 RT Oncogene Res. 1:375-386 (1987).  
 RL [11]  
 RN 21  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RT "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RT factor receptor proto-oncogene RNA synthesis."  
 RT J. Biol. Chem. 266:1746-1753 (1991).  
 RL [12]  
 RN 22  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RT "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene."  
 RT Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924 (1985).  
 RL [13]  
 RN 23  
 RP SEQUENCE OF 540.  
 RX Kohda D.;  
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RL [14]  
 RN 24  
 RP RECEPTOR ACTIVITY.  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Mroczkowski B., Mosig G., Cohen S.;  
 RT "ATP-stimulated interaction between epidermal growth factor receptor  
 RT and supercoiled DNA."  
 RT Nature 309:270-273 (1984).  
 RL [15]  
 RN 25  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Jax I., Kris R., Dombalagian M., Honneger A.M.,  
 RA Howk R., Givol D., Ullrich A., Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.  
 RT Identification of a novel site in EGF receptor."  
 RT J. Biol. Chem. 264:10667-10671 (1989).  
 RL [16]



RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RX MEDLINE=96398132; PubMed=8962717;  
 RA Smith K.D., Davies M.J., Bailey D.V., Hounsell E.F.;  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 RT the epidermal growth factor receptor expressed in Chinese hamster  
 RT ovary fibroblasts.";  
 RL Growth Factors 13:121-132(1996).  
 RN (17)  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; PubMed=10731668;  
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor  
 RT receptor.";  
 RL J. Biochem. 127:65-72(2000).  
 RN (18)  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=98225196; PubMed=9556602;  
 RA Abe Y., Otake M., Inagaki P., Iak I., Schlessinger J., Kohda D.;  
 RT "Disulfide bond structure of human epidermal growth factor receptor.";  
 RL J. Biol. Chem. 273:11150-11157(1998).  
 RN (19)  
 RP REVIEW.  
 RX MEDLINE=87297456; PubMed=3039909;  
 RA Carpenter G.;  
 RT "Receptors for epidermal growth factor and other polypeptide  
 RT mitogens.";  
 RL Annu. Rev. Biochem. 56:881-914(1987).  
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF  
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 CC EGF-like growth factor. Gp30 and vaccinia virus growth factor. Is  
 CC involved in the control of cell growth and differentiation.  
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 CC secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms: 1/p170 (shown here), 2/p60/  
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by  
 CC alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 CC expressed in ovarian cancers.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; X00588; CA25240.1; -;  
 DR EMBL; U95089; AAB53063.1; -;  
 DR EMBL; U48722; AAC50802.1; -;  
 DR EMBL; U48723; AAC50804.1; -;  
 DR EMBL; U48724; AAC50796.1; -;  
 DR EMBL; U48725; AAC50797.1; -;  
 DR EMBL; U48726; AAC50798.1; -;  
 Query Match 45.5%; Score 3111; DB 1; Length 1210;  
 Best Local Similarity 49.5%; Pred.No. 4.3e-15;  
 Matches 625; Conservative 172; Mismatches 364; Indels 102; Gaps 21;

QY ASLSFLDIOIEVOGVLLIAHQVQVPLQRIYRGTLPEEDYALAVLQNGDPLNNTTP 128  
 DB 74 YDLSFLKIQIAGVAVLIALNTVERIPENLIQIRGMYENSLALALNSVD----- 126  
 QY 129 VTGASPGGLRELOLRSLTEILKGVLLIQRPOLCYODTILMKDIPIKRNQALATLIDTNR 188  
 DB 127 --ANKTGKELPVRNLOELHGAVERSNPNALCNVESIQWRDIVSSDFLSLMSMDPQNH 183  
 QY 189 SRACHPSPMCKSKRCWCESEEDCOILTRYCAGGA-RCKGFLPTDCCEQCAAGCTGP 247  
 DB 184 LGSQKCDPSCPNCSGWAGEENCQKLTIKIICAQCSGRGRKSPDSCNHQCAAGCTGP 243  
 QY 248 KHSQDLALHNHNSGICELHCPALVYNTDTFESMPREGRTFGASCVTACPNYLYSTD 307  
 DB 244 RESQDLCKRKRDRATCKDCTPMLNPTTYQMDVNPREGYSGATCVKCKPNYVYTD 303  
 QY 308 VGSCTLYCPHNOVVTMEDGTQRCCKSKPCARCYGLGQYIKANSKFTGITELE-PAG 366  
 DB 304 HGSCVRAAGADSYEM-EDGVKCKCKCEGCRKCNIGIGIEFK-DLSINATNIKHPKN 361  
 QY 367 CKKIFGSLAFPSFSDGDPASNTAPLQPELOVETLEITGYLYISAMPDLSLVSFQ 426  
 DB 362 CTISISDLHLIPVAFRQDSTHTPPLDPELDILKTVKEITGFLIQAMPENRDTHAFE 421  
 QY 427 NLQVIRGRILHNGAVSULTLOGISWGLRSLRELSSGLALIHNTLCEVHTVPMQDLF 486  
 DB 422 NLEIRRTKQHQGFSLAVVSLNITSGLNSLVEISGDVYISGNKMLCVANTINWKLF 481  
 QY 487 RNPQALHTANRPEDCEVBEGLACHQLCARHGCMWGPPTQCVNCSQFLRGQECVECEV 546  
 DB 482 GTSGQKTKIISNRKNSCKATQVCALCPSEGWCPEPRDVCSENVNSGREVCDCNLT 541  
 QY 547 LQGLPREYVNAARHCLPCHPFCOPONGSVTQGEADQVCAYHQPFCVACSPGVK 606  
 DB 542 LGGPRFRFVNSBQICHPBCLPQAMNITLTGSPDNCICAHYIDPHCVKTCPCAVWG 601  
 QY 607 DLSMPWKFPEDEGACQPCPINCTSGSVLDLDDKGPACQASPLTISFNFTVSFLRV 666  
 DB 602 ENNTLT-VMKYADAGHVCHLCPNCTYQCTGPGLEGCTNGPKIP-SIANGVALLLL 658  
 QY 667 PKVSASHLEBRQKRIKYMRLLOETELVEPLTSFGAMPNOAKRIKETELRYKVL 726  
 DB 659 VVALGIGLFMRKRRHIVKRLRLQERLEVEPLTSFGAMPNOALRIKETELFKIKVL 718  
 QY 727 GSGAFGVYKGIWIPDGENYKIPVAIKVIRENTSPKXKEILDEAYVMAGVSPYVRL 786  
 DB 719 GSGAFGVYKGIWIPGEKVKIPVAIKELREASPKANKEILDEAYVMASVDNPHVCRLL 778  
 QY 787 GILTSIVOLVQLMRYGCLLDHVENRGRGLSGQDLINMCQIAKMSYLEDVRLVHRL 846  
 DB 779 GILTSIVOLITLMPFGCLLDYREHKNIQSQYLNNVCQIAKMSYLEDRLLVHRL 838  
 QY 847 AARNVYKSPNHKIKIDFGIARLDIDETEVHADGKVPKXMALESILRRRTTHSDW 906  
 DB 839 AARNVYKTPQHKIKIDFGIARLDIDETEVHADGKVPKXMALESILRRRTTHSDW 898  
 QY 907 SYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMWKCMIMSECR 966  
 DB 899 SYGVTVWELMTFGSKPYDGIIPAREISILSEKGRRLQPICTIDVYIMWKCMIMDADR 958  
 QY 967 PRFRELVEFSRNARBPORFVYIQ-NEDIGPASPLDSTYRSLLEDMDGLVDAAEYLV 1025  
 DB 959 PRFRELITFSKVARBPORYLYIQGERHNLSPDTSNRYRALMDEEDMDVDVDAEYDI 1018  
 QY 1026 PQGPFCDPPAPAGAGGVHRRSSSTRSGGDLTLGLPSEBEARSPLASSEGAGSDV 1085  
 DB 1019 PQGPF-----SSPSTRITLSSLSKTSN- 1043  
 QY 1086 PQGDLGMCAKAGIQTPTDPSPLQRYSDPTVPLPSET-DGYVAPLTCSPQEPYVNP 1143  
 DB 1044 -NSTVACIDRNGLQSPKIKEDSFLQRYSSDPTGALTEDSIDDTFL-----VYPEYING- 1095

QY 1144 DVPFPPSREGFLPAPAPAGATLERAKTLSPGKNGVQVDFAGAVENPEYL-TPQGG 1202  
 Db 1096 SVKCPAGSVQNPVYHNOPLNP-----APSRDPHYGD--PSTAVGNPEYLTVC-- 1143  
 QY 1203 AARQHPHAFSAFENLYYMQ-----DP-----PERGAPSTFGTPTAENP 1246  
 Db 1144 -----PTCVNSTFDSPAHMAQKSHQISLNDPDYQDFPEKAKPNQIFKGS-TAENA 1195  
 QY 1247 EYL 1249  
 Db 1196 EYL 1198  
 RESULT 5  
 EGFR\_MOUSE  
 ID EGFR\_MOUSE STANDARD; PRT; 1210 AA.  
 AC 001279;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112).  
 GN EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RA MEDLINE=93026370; PubMed=1408137;  
 RA Avivi A., Skorecki K., Yayon A., Givol D.;  
 RT "Promoter region of the murine fibroblast growth factor receptor 2  
 (bek/KGFR) gene."  
 RL Oncogene 7:1197-11962(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; and CD-1; TISSUE=Liver, and Decidua;  
 RA MEDLINE=93126380; PubMed=7678348;  
 RA Paria B.C., Das S.K., Andrews G.K., Day S.K.;  
 RT "Expression of the epidermal growth factor receptor gene is regulated  
 in mouse blastocysts during delayed implantation."  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/c; TISSUE=Liver;  
 RA Hibbs M.L.;  
 RT Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B6/C3; TISSUE=Liver;  
 RA MEDLINE=94170986; PubMed=8125255;  
 RA Luetcke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,  
 RA Jenkins N.A., Lee D.C.;  
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF  
 receptor tyrosine kinase."  
 RL Genes Dev. 8:399-413(1994).  
 RN [5]  
 RP SEQUENCE OF 1-714 FROM N.A.  
 RC TISSUE=Brain;  
 RA MEDLINE=9132866; PubMed=2030916;  
 RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;  
 RT "Comparison of EGF receptor sequences as a guide to study the ligand  
 binding site."  
 RL Oncogene 6:673-676(1991).  
 RN [6]  
 RP SEQUENCE OF 969-1117 FROM N.A.  
 RC STRAIN=C3H;  
 RA Eisner D.P., Serrero G.;  
 RT Submitted (Jun-1992) to the EMBL/GenBank/DBJ databases.  
 CC - FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,  
 CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND  
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).  
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC tyrosine phosphate.  
 CC - SUBCELLULAR LOCATION: Type I membrane protein.  
 CC - MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 CC EMBL; X78987; CAA55587.1; -  
 CC EMBL; U03425; AAI17899.1; -  
 CC EMBL; X59698; CAA42219.1; -  
 CC EMBL; L06864; AAS3029.1; -  
 CC EMBL; Z12608; CAA78249.1; -  
 CC HSSP; P11362; PKK.  
 CC MGD; MGI:95294; Egfr.  
 CC InterPro; IPR000494; EGFR\_L domain.  
 CC InterPro; IPR000719; Euk\_Pkinase.  
 CC InterPro; IPR002174; Furin-like.  
 CC InterPro; IPR001245; Tyr\_Pkinase.  
 CC Pfam; PF00069; Pkinase; 1.  
 CC Pfam; PF00757; Furin-like; 1.  
 CC Pfam; PF01030; Recep\_L domain; 2.  
 CC ProDom; PD000001; Euk\_Pkinase; 1.  
 CC SMART; SM00261; FU; 3.  
 CC SMART; SM00219; TYKc; 1.  
 CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 CC PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 CC Transmembrane; Glycoprotein; Receptor; Signal; Transferase;  
 CC Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.  
 CC FT CHAIN 1 24  
 CC FT DOMAIN 25 1210  
 CC FT TRANSMEM 648 670  
 CC FT DOMAIN 671 1210  
 CC FT REPEAT 75 300  
 CC FT REPEAT 390 600  
 CC FT DOMAIN 1028 1071  
 CC FT NP\_BIND 714 981  
 CC FT BINDING 720 728  
 CC FT ACT\_SITE 747 747  
 CC FT DISULFID 839 839  
 CC FT DISULFID 190 199  
 CC FT DISULFID 194 207  
 CC FT DISULFID 215 223  
 CC FT DISULFID 219 231  
 CC FT DISULFID 232 240  
 CC FT DISULFID 236 248  
 CC FT DISULFID 251 260  
 CC FT DISULFID 264 291  
 CC FT DISULFID 295 307  
 CC FT DISULFID 311 326  
 CC FT DISULFID 329 333  
 CC FT DISULFID 332 340  
 CC FT DISULFID 510 523  
 CC FT DISULFID 526 535  
 CC FT DISULFID 539 555  
 CC FT DISULFID 558 571  
 CC FT DISULFID 582 591  
 CC FT DISULFID 595 617  
 CC FT DISULFID 620 628  
 CC FT DISULFID 624 636  
 CC MOD\_RES 680 680  
 CC MOD\_RES 1092 1092  
 CC PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 CC PHOSPHORYLATION (AUTO-) (BY SIMILARITY).

MOD RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1197 1197 PHOSPHORYLATION (AUTO-) (MAJOR SITE)  
 (BY SIMILARITY).  
 FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 19 19 C -> S (IN REF. 2).  
 FT CONFLICT 539 539 C -> W (IN REF. 5).  
 FT CONFLICT 991 991 L -> F (IN REF. 4).  
 FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).  
 SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46D2D2P5 CRC64;

Query Match 45.1%; Score 3081; DB 1; Length 1210;  
 Best Local Similarity 49.3%; Pred. No. 1.6e-155;  
 Matches 627; Conservative 164; Mismatches 371; Indels 110; Gaps 23;

11 LLLALLPAGA--STQVCTGTDMKRLPASPEHLDMLRHLTYQGVQVQGNLELYLPTN 68  
 14 LTLTCAAGALEKKVCGGTNRLTQGTFFEDHLSLQRMNNECVLGNLEIYVQRN 73  
 69 AALSFLQDIQEVQVYLLAHNOVQVPLQRLIYVGTQIFEDNYLAVLUNGDPINNTTP 128  
 74 YDLSFLKTIQEVAVYLLALNTERIPLNLOIIRGNALYENTVALTALISN----- 124  
 129 VTGASPGGLREQLSLTEILKGVLLQBNPOLCQDITLWKDI----FHKNNQALTLI 184  
 125 -YGTNRGTRELPMNLOEILIGAVRPNPILCMNDITQMRDIQVNFMSMSNDL--- 180  
 185 DTNRRAQHPCSPGKSGRCWGESSEDCQLTRTVACGCA-RCKPRLPTDCHQCQAG 243  
 181 -QSHPSCPKCDPSCPNCSGCGEENCQKLTKIKCAQCSHRGSGRSPSCCHQAG 239  
 244 CTGPHSDCLCLHFNHSGIGELHCPALVTYNTDFEEMPNDEGYTGACVYACPNY 303  
 240 CTGPHSDCLVQKQDEATCKDTCPMLNPTTYQMDVNEGYKSGATCVKCCPNY 299  
 304 LSTDVSGCTLVCPHNOEVTADGTQRCSEKSKPCARCYGLGMQYIANKSKFITELE 363  
 300 VTDHSGCVACGPDYEV-EDDGIRKCKCDGPRKVCNGIGIGEFK-DLISINATNIK 357  
 364 -FAGCKTFGSLAFIPESFDGDPASNTAPLOEQLQVETLEITGYIYISAMPDSLDEL 422  
 358 HFKYCTAISGDHLIPVAFKDGSPTRPLDRELEILKTYKEITGFLLIQAMPDNDTDL 417  
 423 SVFQNLQVIRRIILHNGAVSLTLQGLISWLRSLRELGSGLLHNNHTLCVHTVPM 482  
 418 HAFELLEIRGRTHQHGQPSLAVGNTLSLGLSKLISGDVILISNRLVCANTINM 477  
 483 DQFPNPHQALLHTANPEDECVGEGALCHQLCARGHCGWGPPTQVNCQSGFLRGQCE 542  
 478 KKLRTPNQKTKIMNRAEKDCKVNVHNCPLCSSEGCWGPEDPCVSCQVWSGREVE 537  
 543 ECRVLOGLPREVYANRHLCPHQEPQNGSVTQGPBADCVAACHYKDPFVACPS 602  
 538 KNLIEGEPREVENSECIOCHPECLPOANVITCTGRGPDICICAHIDBPHCVKTCPA 597  
 603 GVKPDLASYPIWKEPFDEGACQPCPINCTSCVDLJDKCPAEOASPLTIFNNFTVSF 662  
 598 GIMENNNTL-VMKYADANNVCHLCHANCTYGCAGGAGGQCEVWSPGKIPSIANGIYGL 656  
 663 WLRPKVYASHLKBRQCKIRKYMRELQETELVEPLTBSGAMPNQAQRILKETLRK 722  
 657 LFIYVVALGIGLPMRRRIYKTKLRLQRELVEPLTBSGAMPNQAHRILKETEFKK 716  
 723 VKVLGSAFGTVYKGIWIPDGENVKIPVAILKVLRENTSPKANKEILDEAVYVAGVSPY 782

DL 717 IKVLGSAFGTVYKGIWIPDGENVKIPVAILKVLRENTSPKANKEILDEAVYVAGVSPY 776  
 783 SRLGICLSTVQVLTQMLMPYGCCLLDHRENRGRIGSODLNMCMQIAKASYEDVLY 842  
 777 CRLLGICLSTVQVLTQMLMPYGCCLLDHRENRGRIGSODLNMCMQIAKASYEDVLY 836  
 843 HRDLAARVAVLKYSPHNYKXITDFGLARLLDIDETRYHADGCKVPIKMALESILRRFTQ 902  
 837 HRDLAARVAVLKYSPHNYKXITDFGLARLLDIDETRYHADGCKVPIKMALESILRRFTQ 896  
 903 SDWSYGVTVWELMTFPAKPYDGI PAREIDPLEKGERLPQPPICITIDVYIMYKCMID 962  
 897 SDWSYGVTVWELMTFPAKPYDGI PAREIDPLEKGERLPQPPICITIDVYIMYKCMID 956  
 963 SECRPFRELYSEFSRMAPDPORFVVIQ-NEDLPASPLDSTFRSLLEDMDMDLVDAE 1021  
 957 ADSRPFRELYSEFSRMAPDPORFVVIQ-NEDLPASPLDSTFRSLLEDMDMDLVDAE 1016  
 1022 EYLVPOQGFPPDPAFGAGMTHRRSSSTRSGGDLTLGLEPSEEDAPRPLAPSEGA 1081  
 1017 EYLVPOQGFPPDPAFGAGMTHRRSSSTRSGGDLTLGLEPSEEDAPRPLAPSEGA 1042  
 1082 GSDVFDGDLGKAAGLQSLPTHDPSLQYSEDPVPLPSET--DGTVAPLTCSPQPEY 1139  
 1043 TSN-----NSTYACINRNGSCRWKEDATLQYSSDPTAVEDNIDDAFL-----PVPEY 1092  
 1140 VNQDPVAPQPPSPREGPLPARPAGATLBERAKTLSPKNGVYKDFVAFGAVENPEYL-T 1198  
 1093 VNQDPVAPQPPSPREGPLPARPAGATLBERAKTLSPKNGVYKDFVAFGAVENPEYL-T 1141  
 1199 POGGAAPQPPSPREGPLPARPAGATLBERAKTLSPKNGVYKDFVAFGAVENPEYL-T 1242  
 1142 AQ-----PCTLSSGFSPALMTIQGSHQSLMDPDYQDFFPKETYPNGIFKG-PT 1191  
 1243 AENPEYGLDVP 1254  
 1192 AENPEYGLDVP 1203  
 RESULT 6  
 ID ERB4 HUMAN STANDARD; PRT; 1308 AA.  
 AC 015303;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-4 precursor (BC 2.7.1.112)  
 DE (p180erbB4) (Tyrosine kinase-type cell surface receptor HER4).  
 GN ERB4 OR HER4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN 11  
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=93189574; PubMed=8383326;  
 RA Plozman G.D., Cuioscu J.-W., Whitney G.S., Green J.M., Carlton G.W.,  
 RA Foy L., Neudauer M.G., Shoyab M.,  
 RA "ligand-specific activation of HER4/p180erbB4, a fourth member of the  
 RT epidermal growth factor receptor family".  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750 (1993).  
 [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=97476287; PubMed=9334263;  
 RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plozman G.D.,  
 RA Klagsbrun M.;  
 RA "A novel transmembrane domain isoform of HER4/erbB4, isoform-specific  
 RT tissue distribution and differential processing in response to  
 RT phorbol ester".  
 RL J. Biol. Chem. 272:26761-26768 (1997).

CC - FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NR-  
 CC 2, NR-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND  
 CC NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.  
 CC NOT ACTIVATED BY EGF, TGF- $\alpha$ , AND AMPHIREGULIN.  
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC - SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB  
 CC RECEPTORS (POTENTIAL).  
 CC - SUBCELLULAR LOCATION: Type I membrane protein.  
 CC - ALTERNATIVE PRODUCTS: 2 ISOFORMS, JM-A (SHOWN HERE) AND JM-B;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER  
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED  
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND  
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN  
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.  
 CC - TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,  
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,  
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,  
 CC LUNG, SALIVARY GLAND, AND PANCREAS.  
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC EMBL: L07868; AAB59446.1; -  
 CC HSSP: P1362; 1FGK.  
 CC Gene: HGNC:3432; ERBB4.  
 CC MIM: 600543; -  
 CC InterPro: IPR000494; EGFR\_L\_domain.  
 CC InterPro: IPR000719; Euk\_pkinase.  
 CC InterPro: IPR002174; Furin-like.  
 CC InterPro: IPR01245; Tyr\_pkinase.  
 CC InterPro: IPR004019; YLP\_motif.  
 CC Pfam: PF00069; pkinase; 1.  
 CC Pfam: PF00757; Furin-like; 1.  
 CC Pfam: PF01030; Recep\_L\_domain; 2.  
 CC Pfam: PF02757; YLP\_2\_domain; 1.  
 CC ProDom: PD000001; Euk\_pkinase; 1.  
 CC SMART: SM00261; FU; 4.  
 CC SMART: SM00219; TYKc; 1.  
 CC PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC Transmembrane, Glycoprotein, Multisubunit family; Receptor; Signal;  
 CC Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 CC Alternative splicing.  
 CC  
 CC FT CHAIN 1 25 POTENTIAL.  
 CC FT DOMAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.  
 CC FT TRANSMEM 652 651 EXTRACELLULAR (POTENTIAL).  
 CC FT DOMAIN 676 1308 POTENTIAL.  
 CC FT DOMAIN 186 1308 CYTOSOLASMIC (POTENTIAL).  
 CC FT DOMAIN 496 633 CYS-RICH.  
 CC FT DOMAIN 718 985 PROTEIN KINASE.  
 CC FT NP\_BIND 724 732 ATP (BY SIMILARITY).  
 CC FT BINDING 751 751 ATP (BY SIMILARITY).  
 CC FT ACT\_SITE 843 843 BY SIMILARITY.  
 CC FT DISULFID 189 197 BY SIMILARITY.  
 CC FT DISULFID 193 205 BY SIMILARITY.  
 CC FT DISULFID 213 221 BY SIMILARITY.  
 CC FT DISULFID 217 229 BY SIMILARITY.  
 CC FT DISULFID 230 238 BY SIMILARITY.  
 CC FT DISULFID 234 246 BY SIMILARITY.  
 CC FT DISULFID 249 258 BY SIMILARITY.  
 CC FT DISULFID 262 289 BY SIMILARITY.  
 CC FT DISULFID 293 304 BY SIMILARITY.

FT DISULFID 308 323 BY SIMILARITY.  
 FT DISULFID 326 330 BY SIMILARITY.  
 FT DISULFID 503 512 BY SIMILARITY.  
 FT DISULFID 507 520 BY SIMILARITY.  
 FT DISULFID 523 532 BY SIMILARITY.  
 FT DISULFID 536 552 BY SIMILARITY.  
 FT DISULFID 555 569 BY SIMILARITY.  
 FT DISULFID 559 577 BY SIMILARITY.  
 FT DISULFID 580 589 BY SIMILARITY.  
 FT DISULFID 593 614 BY SIMILARITY.  
 FT DISULFID 617 625 BY SIMILARITY.  
 FT DISULFID 621 633 BY SIMILARITY.  
 FT MOD\_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VASPLIC 626 648 N-ISOFORM JM-B).  
 SQ SEQUENCE 1308 AA; 146807 MW; 58AEB0965D86761 CRC64;  
 Query Match 42.7%; Score 2920.5; DB 1; Length 1308;  
 Best Local Similarity 44.6%; Pred. No. 5,4e-147;  
 Matches 604; Conservative 178; Mismatches 384; Indels 187; Gaps 26;  
 QY 9 MGLLLALLPGAA-----STQVCTGTOMKALRLPSFTHDMLHLVGGCCVQGNLELY 64  
 DB 8 WWWWSTLVAAAGTVPDSQSVACGTENKLSLSLDEQYALRLYENCCEVWGNLEITS 67  
 65 LPTNASTSFLDIDIEGVYVLIANQVRQVPLRLIVRSTOLFEDNYVALAVDNGDPLN 124  
 DB 68 IEHRDLSFLRSVREYVGYVVALNQRVPLELRLIRIGKLYEDRYALALFLNVRKDG 127  
 QY 125 NTTVVTGASREGLELELRSITELKSGVLIQRNPOLCYODTILMKOIFKNNOLATLI 184  
 DB 128 NF-----GLQELGLKXLTTELINGVYVDQKFLCYADITHMODIVANPWPSNLTIV 178  
 QY 185 DTNRSRACHPSPWCKGSRGCGESSDCCSLTRTVACAGC-ARCKGPLPTDCHGQCAAG 243  
 DB 179 STNSSGCGRCHSKCTG-RCKGPTENHCOTLITRVACQDDRCYGVYVDDCCHREACAG 237  
 QY 244 CTGPKHSDDCLAHFNHSGICEHLCPALVTYNTDTFESMPNPEGRTYFGASCYTACPNY 303  
 DB 238 CSGPKDNDCEFCANFNDSGACVTCQPTQFYVNPFTPLEHNFNAKYTGAFVCKKCPHNF 297  
 QY 304 LSTVSGCTLVCPPLHNEVLAEDGTQCEKSCKCAVCGIGMQYIKANSKFIIGLELE 363  
 DB 298 V-VDSSCVRAKCPSSKREV-BENGIKMKCKPCTDIPCACGIGISMSQYVDSNDK 355  
 QY 364 FAGCKKIFGSLAPFESFDPPASNTAPLPQPOLQVETLEITGYLYISAMPDSDLDS 423  
 DB 356 FINCTKINGNLIPLVTGHHGDPYVAIDPEKLNVRVREITGFLINISWPNMDFG 415  
 QY 424 VFQNLQVIRGILNGAVSLTQLGSLGSLRLSGLALIHNTHLCPHYTPMD 483  
 DB 416 VFSNLVTIGGVVLSGLLILKQGLTSLQFSLKELISAGNIYITDNSLCLYHTTNWT 475  
 QY 484 QLFNPPQALLHTANRPEDECVGEGLAHQHLCARGHCGGPTQCVNCSOFLRGOECVEE 543  
 DB 476 TLFTTINQRIVIRNRKAEKNTABGWCNHLCSDDGWMGPGPOGLSCRRFSRGRILIES 535  
 QY 544 CRVIGGLPREYVNAHRLCPHPCOP-QNGSVTCFGEADQCCVACAHYKPPFVACPS 602

Db 536 CNYDGFREFRNGSTVECDPCCEKEDGLTCHGPGDNTCKSFKFGKPCVCEKCPD 595  
 Qy 603 GVXEDLSYPMWKEPDEGACQCPINCSTHSCVDLDDKGPAPGASPLTSIF-----NNF 658  
 Db 596 GLOQANSF--IFKYADDRECHPCPNCTGCGNCPSTSHDC-----IYPMWTHS 642  
 Qy 659 TVSWMLVPRKVSAS-----HLEKRRQOKIKYIMRLLOTELVLPPTS 703  
 Db 643 TLPGHARTPLIAAGVIGLPLIVIGLTFVAVYRRKSIKKRRALRRL-ETELVLPPTS 701  
 Qy 704 GAMPNQAMRILKETELRKVVGSGAFGYKGIWIPDENVKIPAIIVLENTSPKA 763  
 Db 702 GTANQAOQLRLKETELRKVVGSGAFGYKGIWIPDEETVKIPAIIVLENTSPKA 761  
 Qy 764 NKEILDAVYVAGVGSPPVSRRLGICLTSTVQVLTQMPYGLLDHVRNGRLSGDLL 823  
 Db 762 NVEEMDEALIMASNDHRLVRLGLVCSPITQVLTQMPYGLLDHVRNGRLSGDLL 821  
 Qy 824 NMCMQIAKMSYLEDVRLVRLDLAARVAVKSPVHVKITPGLARLLEDETEYHADGK 883  
 Db 822 NMCMQIAKMSYLEDVRLVRLDLAARVAVKSPVHVKITPGLARLLEDETEYHADGK 881  
 Qy 884 VPIYMALESILRRFTHQSDVMSYGVTVMLTFGAKPYDGIPIAREIPDLLEKGERLPQ 943  
 Db 882 MPKMALECIHYKFTHQSDVMSYGVTVMLTFGAKPYDGIPIAREIPDLLEKGERLPQ 941  
 Qy 944 PPICTIVMIMVCMIMIDSECRPRELVSFSPMRADQRFVYVIONED-LGAPSLDS 1002  
 Db 942 PPICTIVMIMVCMIMIDSECRPRELVSFSPMRADQRFVYVIONED-LGAPSLDS 1001  
 Qy 1003 TFRSLDEDDMGDLVAEEYLVPOQGFPCPDPAAGAGVHHHRSSSTRSGGDLTLG 1062  
 Db 1002 KFFGNLDEEDLEMMDAEYLV-QAFNIPPP-----YTSRARIIDSRS-----ETG 1049  
 Qy 1063 LEPSEBAPPS-----PLAP-SEGAGSDVFDGIGWG 1093  
 Db 1050 HSPRPAYTPWNGQFYVRDGFAGAGVSVYRAPSTIIEBAPVACATLEIPDDSCNG 1109  
 Qy 1094 AAKLOSLPTHDSPLORYSEDPVPLPS-----ETDGVAVPLTCSPOPEVNOQDVA 1146  
 Db 1110 TLKRPVAPHVQSDSTORYSADPTVFAPERSPGELDEEGVMPMDKPKQOYLNVYE-- 1167  
 Qy 1147 PQPSPREGPLPAARPGATLERAKTLSPKNGVYKDVFAFGAVENPEYLLTPQGAAPQ 1206  
 Db 1168 -----ENPFVSR-----KNGDLQ-----ALDNPEYHNASNG--- 1194  
 Qy 1207 PHPPPA-----FSPAFDLVYWDQDPPEBGA 1232  
 Db 1195 --PFAKDEYVNEPLVNTANTLGAKEYVKNMILSPERAKKAFDNPDIWNHSLPFRST 1252  
 Qy 1233 --PPSTFKGTP-----AENPEYL 1249  
 Db 1253 LQHPDYLGEXSTKYFYKQNGRIRIYANPEYL 1285

RA Marchionni M.A., Kelly R.A.;  
 RT "Neuregulin promotes survival and growth of cardiac myocytes."  
 RT Persistence of ErbB2 and ErbB4 expression in neonatal and adult  
 RT ventricular myocytes."  
 RL J. Biol. Chem. 273:10261-10269(1998).  
 RN [2]  
 RP SEQUENCE OF 848-901 FROM N.A.  
 RC TISSUE=Sciatic nerve;  
 RX MEDLINE=9122560; PubMed=2025425;  
 RA Lai C., Lemke G.;  
 RT "An expanded family of protein-tyrosine kinase genes differentially  
 RT expressed in the vertebrate nervous system.";  
 RL Neuron 6:691-704(1991).  
 RN [3]  
 RP SEQUENCE OF 1031-1198 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;  
 RX MEDLINE=97184212; PubMed=9030624;  
 RA Carroll S.L., Miller M., Frohner P.W., Kim S.S., Corbett J.A.;  
 RT "Expression of neuregulin and their putative receptors, ErbB2 and  
 RT ErbB3, is induced during Wallerian degeneration.";  
 RL J. Neurosci. 17:1642-1659(1997).  
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN. NRG-  
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND  
 CC NOT ACTIVATED BY EGF, TGF- $\alpha$ , AND AMPHIREGULIN (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB  
 CC RECEPTORS (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING  
 CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS  
 CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE  
 CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND  
 CC HEART.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC EMBL: AF041838; AAC08899.1; -  
 CC HSPB: U52531; AAC53051.1; -  
 DR HSPB: P11362; IFCR.  
 DR InterPro: IPR000494; EGFR\_L domain.  
 DR InterPro: IPR000719; Furin-like.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR004019; YLP motif.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF01030; Recep\_L domain; 2.  
 DR Pfam: PF02757; YLP\_2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00261; FU; 4.  
 DR SMART: SM00219; Tyrc; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Tyrosinase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1308  
 FT DOMAIN 26 651  
 FT TRANSMEM 652 675  
 FT DOMAIN 676 1308  
 FT CYTOPLASMIC (POTENTIAL).

```

FT DOMAIN 186 334 CYS-RICH.
FT DOMAIN 496 633 CYS-RICH.
FT NP BIND 718 985 PROTEIN KINASE.
FT BINDING 724 732 ATP (BY SIMILARITY).
FT ACT_SITE 751 751 ATP (BY SIMILARITY).
FT DISULFID 843 843 BY SIMILARITY.
FT DISULFID 189 197 BY SIMILARITY.
FT DISULFID 193 205 BY SIMILARITY.
FT DISULFID 213 221 BY SIMILARITY.
FT DISULFID 217 229 BY SIMILARITY.
FT DISULFID 230 238 BY SIMILARITY.
FT DISULFID 234 246 BY SIMILARITY.
FT DISULFID 249 258 BY SIMILARITY.
FT DISULFID 262 289 BY SIMILARITY.
FT DISULFID 293 304 BY SIMILARITY.
FT DISULFID 308 323 BY SIMILARITY.
FT DISULFID 326 330 BY SIMILARITY.
FT DISULFID 503 512 BY SIMILARITY.
FT DISULFID 507 520 BY SIMILARITY.
FT DISULFID 523 532 BY SIMILARITY.
FT DISULFID 536 552 BY SIMILARITY.
FT DISULFID 555 569 BY SIMILARITY.
FT DISULFID 559 577 BY SIMILARITY.
FT DISULFID 580 589 BY SIMILARITY.
FT DISULFID 593 614 BY SIMILARITY.
FT DISULFID 617 625 BY SIMILARITY.
FT DISULFID 621 633 BY SIMILARITY.
FT MOD_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT CARBOHYD 138 138 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 174 174 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 253 253 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 358 358 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 410 410 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 473 473 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 495 495 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 548 548 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 576 576 N-LINKED (GLCNAC) (POTENTIAL).
FT CARBOHYD 620 620 N-LINKED (GLCNAC) (POTENTIAL).
FT CONFLICT 1062 1062 S -> N (IN REF. 3).
FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).
SQ SEQUENCE 1308 AA: 146957 MW: D944BB0996A08B41 CRC64;

```

Query Match 42.5%; Score 2906; DB 1; Length 1308;  
 Best Local Similarity 44.5%; Pred. No. 3.1e-146;  
 Matches 603; Conservative 185; Mismatches 392; Indels 174; Gaps 26;

```

QY 1 MELA-ALCRWGLL--ALLPPGAATOVCTGTDMLRLPASPETHLMDLRHYOGCOVQ 57
DB 1 MKLATGLWVGSLLVAVARTVQPSASOSVCGTENKLSLSLDEQVRLRKYYENCEVYM 60
QY 58 GNLELYLPTNALSFLQDIQEVQGYVLIHNVORVPLQRLIVRGTOLEFEDNVALAVL 117
DB 61 GNLEITSIEHNRDLSPRSIREVTGYVLIHNVORVPLQRLIVRGTOLEFEDNVALAVL 120
QY 118 DNGDPLNNTTPTVAGSPGRLREQLRSLTEILKGVLIORNPOLCVQDITLWMDIHKXN 177
DB 121 LNRKXDNF-----GLQELGKLKLTETILNGSVYVDONKFLCADITIHMODIVRNM 171
QY 178 QLAULTLIDNRSACHPCSPMGKSRGWESSEDCSLTRTVAGGC-ARCKGFLPTDCC 236
DB 172 PSNMTLVSTIGSGCGCHKSCG-RCWGPTEHNCQTLFRTVCAEQCDRCRGVYVSDCC 230
QY 237 HECGAGCGSPKSDCLACHFNHSGICELHCPALVTYNTDTEESNPNBGRYTFASCV 296
DB 231 HRECGAGCGSPKSDCLACHFNHSGICELHCPALVTYNTDTEESNPNBGRYTFASCV 290
QY 297 TACPYNLTSDVSCCTLVCPILNQEVTAEDGTORCEKSPCARVCYGLGMOYIKANSKF 356
DB 291 KKCPHNFV-VDSSSCVACAPSSKMEV-EEGIMCKRCPTDIDCKACDGIQTGSLMSAQTV 348

```

```

QY 357 IGITELFAGCKKIIFGSLAFPLESPFDGDPASNTAPLOPEQLQVETLEEITGYLYISAMP 416
DB 349 DSNINDKINCTKINGNIIIFLVTHGHPYVAIDIDPEKLNVERTYREITGFLINQWMP 408
QY 417 DSLPLSTFQULVIRGILHNGAVSLTQSLGISWGLRSLRELSSALIHNTNLCF 476
DB 409 PNMTDFSVFSLVYIGGVLSGLSLILKQGGITSLQFOSLKEISANIVITDONSICY 468
QY 477 VHTVMDLFPNPHQALHTANPEDECEVSGLAGCHQACARHCGMPPTQCVCNCSQFLR 536
DB 469 YHTIMWTLLFSTVNDRIYIRDNRAENCTAAGMCMNLHCSNDGCMGPRDCLSCRRSR 528
QY 537 GQECVEECVLIQGLPREVNAHCLPCHPECOP-QNSVTCFGEADQVCAHAYKDPF 595
DB 529 GKICIESCNLYDEGEREFENSGICEVCSQCEKMEGDLTCHGCPDCTCKSHFKDQPN 588
QY 596 CVARPPSVKRDLSMPYMKPEPDEGACQCPINCTHSCVLDKCGCAEGRASPLTIF 655
DB 589 CVERCEPDVQANNSF--IFKYADODRECHPCHPCTQCGNCPSTSHDC-----IY 635
QY 656 ---NFTVSEWLRVPKVSAS-----HLEKROCKIRKYTRRLLOETEL 696
DB 636 YPMTHGSHLPQHARTPLIAAGVIGGLFVLMALTFAYVYARKSIKKRALRFL-ETEL 694
QY 657 VEPLTPSGAMPNQAOMRLKETEIRKYKVLGSAFGTYKGIWIPDGENVKIIPAIAKTLR 756
DB 695 VEPLTPSGAMPNQAOMRLKETEIRKYKVLGSAFGTYKGIWIPDGENVKIIPAIAKTLN 754
QY 757 ENTSPKAKEILDEAVYVAGSPVSVSLGICTSTVQVLTQVLTQVLTQVLTQVLTQVLTQ 816
DB 755 ETTGKXANVEEMDELIAASVDHPLVRLGVCSPITQVLTQVLTQVLTQVLTQVLTQVLTQ 814
QY 817 LGSQDLNWCQIAKMSYLEDVRLVHRDLAARVLYKSPNHVKITDGLARLLDIDETE 876
DB 815 IGSQDLNWCQIAKMSYLEDVRLVHRDLAARVLYKSPNHVKITDGLARLLDIDETE 874
QY 877 YHADGKXPIKMMALLESILRRFTHOSVMSYGVTVVETLFFGAKPYGIPAREIPDLE 936
DB 875 YHADGKXPIKMMALLESILRRFTHOSVMSYGVTVVETLFFGAKPYGIPAREIPDLE 934
QY 937 KGERLPQPICTIDVYIMVACMMIDSECRPRELSEFSRMAADPORFVIONED-LG 995
DB 935 KGERLPQPICTIDVYIMVACMMIDSECRPRELSEFSRMAADPORFVIONED-LG 994
QY 996 PASPLDSTFTYSLLEDMDGLVDAEELVYQOGFPCDP----- 1035
DB 995 LPSPDSTFTYSLLEDMDGLVDAEELVYQOGFPCDP----- 1033
QY 1036 ---APGAGMTHRRHRSSTRSGGDLTLGLEPSEEEAPRSLAPSEGAGSDVFGDGLM 1092
DB 1054 PAYTMSSGCVFYQDGFATQOQ---KMPYTAITSTIPAPVA--QATLMDMDSDCN 1108
QY 1093 GAAKGLQSLPTHDSPLQRYSEDPVLPJS-----ETDGYVAPLTCSPQPEYVNOQDV 1145
DB 1109 GLTRKFPVPHVQEOSSTQRYGADPTVFAPERNPAELDEBQYMTPMHDKPQOEYLNPEV- 1167
QY 1146 RPQPSRREGPLAPRPAAGTLERAKTSLSPKNGVYKQVAFGCAVENPEVLTQGGAP 1205
DB 1168 ---ENPFVSR-----KNGDLQ-----ALDPEYHSSSG-- 1194
QY 1206 QPHPPA-----FSPAFLNLYWDDPBERG 1231
DB 1195 ---PEKADDEVNEPLVNTFTNALGNAEYKNSLSLVSPEKAKKAFQMDPVWNSLSPRS 1251
QY 1232 A--PPSTKGTPT-----AENEYL 1249
DB 1252 TLQHPDYLOEYSTKYFYKONGRIIRPIYANENEYL 1285

```

RESULT 8  
 XMXX XIPIWA  
 ID XMXX XIPIWA  
 AC P13388;  
 STANDARD; PRT; 1167 AA.







```

FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 286 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 566 573 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPIC 141 183 ELISGVYIEKNDKLCMDTIDMDRIDVDRDAEIVYKONGR
SC -> GCFPMVPSGLTPQADWYLLDDDDPRLLTSLASSK
VPYTLAAV (IN SHORT ISOFORM).
FT VARSPIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 149097 MW; 7201E7F66CA374BD CRC64;

```

Query Match 35.1%; Score 2398.5; DB 1; Length 1342;  
 Best Local Similarity 40.0%; Pred. No. 2e-119;  
 Matches 527; Conservative 192; Mismatches 458; Indels 139; Gaps 33;

```

QY 10 GLLALLPFGAA--STQVCTGTDKRLPASPEHLDMKRLHYQGGQVQGNLELYLPT 67
DB 11 GLLFSLARGSEVNSQAVCPGLTNGLSVGDAMENOYQTLKYKXEREEVVMGNLEIVLTS 70
QY 68 NASLSFLDIOIEVQGVLLAHNOVRQVPLQRLRIYRGTOLEFEDNALAVLNDGDDLTNT 127
DB 71 NADLSFLQIREVTGYLVANMEFSLPLPRLRVAGTQVQGGKFIIVM-----LNYNT 125
QY 128 PVTGASPGGLRELQRLSTLEILKGVLLIQRPOLCYODTILWKDIFHKNNQALTLIDTN 187
DB 126 ---NSSHALRQLRLTQLTEILSGVYIEKNDKLCMDTIDMDRIDVDRD---AEIVKD 178
QY 188 RSARCHPCSPMCKSGSKWESSEDCSLRRTYCAGGC-ARKGRLPTDCHGQCAAGCTG 246
DB 179 NGSCSPCHREVCKG-RCMGPGSEDCQTLTKTICAPQNGCHCGPMPNQCDECAAGGSG 237
QY 247 PKHSDCLAFHFNHSGICELHCPALTYNTDTFESMPNBEGRYTFGASCTACPNYLTST 306
DB 238 PQDTDFACRHHNDSACVPRGCPGLVYMKLTFGLEPNNHTYXQVGVAVASCPNFFV-V 296
QY 307 DVGSGTCLVPLHNOETADGTQRCCKSKPCARVCYGYGMQYIANSKF--IGTTELE- 363
DB 297 DQTSQVRACPRDKMEVD-RNGDKMCEPCGGLCPKACEGG-----SGSHFQTVDSNDIG 350
QY 364 FAGCKKIFSLAFLPESFDGDPASNTAPLQBPOLQVFTLEETIGTVLSAPRDLPLDS 423
DB 351 FVNTCTKILGNDPLTILGNDGPMHKIPALDPEKLVFRVRIITGLYLIQSPMPMHNS 410
QY 424 VFQNTQVIRGRILHNGAVS-LTLQGLISWIGLSRLRELSSGLAIHNNTHLCFVHTVPM 482
DB 411 VESNLTITIGRSLYNNGFSLLIMKNLNTVSLGFRSLKETSAGRIYISANRQCLCHNSLNM 470

```

```

QY 463 DQLEPNPQALLHTA-NREDECVGEGLACHQLCARHCKGPRPTQCVNGSQFLRGQECV 541
DB 471 TKVIRGPTEERLDITKHNRPRRDCAEGKVCPLCSSGGCWPGGQCLCSHNSRGVCV 530
QY 542 EECRYTQGLPREVYNAHCLPCHPECOPONGSVTCFGEADQCVAAHYKDPFCVAPCP 601
DB 531 THCNFLNGEPPEFAHEACFCSECPDECMESTATONSGSDTCAQCAHFRGPCVSSCP 590
QY 602 SGVPRDLSYMEIMKRPDEGACOPCPINCTHSCVLDLDDKGPAP-----QRA SPLTS 654
DB 591 HGVLG--AKGIYKVPDVQNECRPCHEKCTGCKGPELDQCLGQTLVLIGKTHLTALTV 648
QY 655 FNNFVTSWKLAVPKVSASHLEKROQKRTKTRRLLOETLVEPLTSSGAMPQAOQRI 714
DB 649 IAGLVVIFM-----LQGTFLYRRRRRIQNKAMRYLERGESIEPLPS-ERAKKVLARI 703
QY 715 LKETELRKVLKLSGAFGTYYKGIPIPGENVKIPIVAKIVRENTSPRANKELIDEAVM 774
DB 704 FKETELRKVLKLSGVTGTVAHKGWIPGESIKIPVICIIVIEDSKGSGPQAVTDHMLAI 763
QY 775 AGVSPVYSRLIGTISTVOLVTQLMPIYGLDHDVENRGLSGQDILLNMQOIAKMS 834
DB 764 GSLDHAHIVRLIGCPGSSLOLVTVQYPLGSLDHDVHQRGALPQLLNGVQIAKMY 823
QY 835 YLEDRVLRDLAARNVLVKS PNHVKITPGLRLDIDETEHADGKVPKIMMALESI 894
DB 824 YLEHGMVHRLAARNVLKSPSOVQVADFEVALLPDDKQLLYSEAKTIPKMALESI 883
QY 895 LRRFTHQSDVSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICITDVMI 954
DB 884 HFQKTHQSDVSYGVTVWELMTFGAEYAGRLAEVVDLLEKGERLAQPOICITDVMI 943
QY 955 MYKCMNIDSEKRPFRFELVSFSRMARDPQAFVYIQNEDLCPA---SPLDSTFYSLLIED 1011
DB 944 MYKCMNIDENIRPFEKELANFTPMADPPRYLVIKRES-GGIAPGEPPEGLTNKXLEE 1002
QY 1012 DDMGDLVDAEYVLPQCGFPCDPAPAGAGVHHRHSSSTRSGGDTLGLPE-SEEA 1070
DB 1003 VELEPELDLDDLDAEE-----NLTATTIGSALSLVGLNRRPG 1043
QY 1071 PRSPLAPSEGAGSDVFDGLMGAKGIQSLPTHD-PSPLQRYSEDPVLP-----SE 1123
DB 1044 SQSLSPSSGY-MPMNQNLGSCQESAVSSSRCPVSLH-----PMRGLCLASES 1096
QY 1124 TDGVYA-----PLTCSQPE-----YNNQPVNRQOPSREGP----- 1156
DB 1097 SEGHTGSEALQEKVSMCRSRSRSPRPGDSAYHSQRHSLTPTVPLSPGLUEEDV 1156
QY 1157 ---LPAAPAGATLERAKTLP--GRNGV-----KDVFAFGAVERPEYLTPOGGAAP 1205
DB 1157 NGYMPDTHLKGTSRREGTSSVGLSGLTEBEDD-----EEYETMKRRRRSP 1208
QY 1206 QHPPPAPSPAFDNLVYWD-----QDPPEGAPSTKGTPTANPEYL 1249
DB 1209 -PHPPPRPSLLELGEYWDVGSLSVSTQSCPLHPVPIPTAGTTPDEDEYEM 1263

```

RESULT 10  
 ERB3 RAT STANDARD; PRT; 1339 AA.  
 AC Q62799; Q62955;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erdb-3 precursor (EC 2.7.1.112)  
 DE (C=erdb3).  
 GN ERB3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 CX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.

CC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=6096535; PubMed=8522190;  
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;  
 RT "Cloning of the rat ErbB3 cDNA and characterization of the  
 RL recombinant protein.";  
 RN Gene 165:279-284 (1995).  
 RN [2]  
 RP REVISIONS TO 85; 513 AND 565.  
 RA Hellyer N.J., Koland J.G.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 922-1097 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
 RX MEDLINE=97184212; PubMed=9030624;  
 RA Carroll S.L., Miller M.L., Frohert P.W., Kim S.S., Corbett J.A.;  
 RT "Expression of neurogulin and their putative receptors, Erb2 and  
 Erb3, is induced during Wallerian degeneration.";  
 RL J. Neurosci. 17:1642-1659 (1997).  
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTRK.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
 CC -1- PFM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sb-sib.ch](mailto:license@sb-sib.ch)).

FT	DISULFID	231	243	BY SIMILARITY.
FT	DISULFID	246	255	BY SIMILARITY.
FT	DISULFID	259	286	BY SIMILARITY.
FT	DISULFID	290	301	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	323	327	BY SIMILARITY.
FT	DISULFID	500	509	BY SIMILARITY.
FT	DISULFID	504	517	BY SIMILARITY.
FT	DISULFID	520	529	BY SIMILARITY.
FT	DISULFID	533	549	BY SIMILARITY.
FT	DISULFID	556	573	BY SIMILARITY.
FT	DISULFID	576	585	BY SIMILARITY.
FT	DISULFID	589	610	BY SIMILARITY.
FT	DISULFID	613	621	BY SIMILARITY.
FT	DISULFID	617	629	BY SIMILARITY.
FT	CARBOHYD	126	126	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	250	250	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	408	408	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	414	414	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	437	437	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	469	469	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	522	522	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	566	566	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	616	616	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CONFLICT	1028	1028	L -> P (IN REF. 3).
SEQ	SEQUENCE	1339 AA; 147545 MW; 0AA5F2402BPF01E CRC64;		

  

Query Match	34.0%;	Score 2325.5;	DB 1;	Length 1339;
Best Local Similarity	40.5%;	Pred. No. 1.5e-115;		
Matches 521;	Conservative 169;	Mismatches 436;	Indels 159;	Gaps 34;

  

QY	3	LAALCRWGLLALLPPGAA---STQVCTGTDMLRLPASFTLDMRLHYOGQVVOGN	59
DB	7	LQVLC---FLSLARQSEMNSQAVCGTUNGSLVGDADNOYQTVLYKECEVWGN	62
QY	60	LELTPLTNASLFDIOQEOGVYLLAHNVQRVPLQRIIVGTOLEFNVALAVADN	119
DB	63	LEIVTGNADLSFQWIREVTGYLVANMFSSVLPNLELVVGTOYVQDKFATFVN--	120
QY	120	GDPLNNTTPVTGASPGGIRELQSLRSLTEILKGVLIGRNPOLCYODTILMDIFRKNQUL	179
DB	121	---LNYNT---NSSHALRQLKPLQVLEILSGVYIEKNDLCHMDITDMRDIVRV--	170
QY	180	ALTLDITRSRACHPCSPCKGSRCKGSESSDQSLRTVAGGC-ARCKPLPTDCHE	238
DB	171	GAEIVKKNKANCPCEHVECKG-RQWGPDPDCCIITKTICAPQNGRCFQPNQOCHD	229
QY	239	OCAGCTGPKKSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTA	298
DB	230	ECAGGSGSPQDTCFACRFRFDSGACVRCGEPLVYNKLTQLBPNHTKXYGVGVYAS	289
QY	299	CPNYLSTDVSSCTLVCPLNQOEVTAEEDGTORCEKSKPCARVYGL--GMQYIKANSKF	356
DB	290	CPNHFV-VDQTFVCYACAPPDQMEVD-KHGLKCEPCGGLCPKACEGSGSGRYQTVDSN	347
QY	357	IGITELFAGKKIFGSLAFPLPESFDGPPASNTAPLOEOLQVETLEITGYLYISMP	416
DB	348	ID---GVNCTKILANIDFLITGLVNDPMHKITALDBEKLVNRTVREITGYLYISMP	403
QY	417	DSLPLSVFONLQVIRGRILHNGAVS-LTLOGIGISWGLRSLRELSGLALIHNTLHC	475
DB	404	PHMNFVSFSLITIGRSLYNGFSLIMNLTWSLIGFSLKXISAGRYISANQOLC	463
QY	476	FVHTVPWQLFRNHQALHTA-NRPEECYGBELACHQLCARHCHGPGFTQCVNSQSF	534
DB	464	YHSHSNMTRLLRGSSEERLDIKYRPLGECCLAEKVCDDPCSSGSGCWPGGQLSCRY	523
QY	535	LRGQECVVEECVLOGLPREYVNAHCLPCHPECOPONGSVTCFPEADQVCACAHYDPP	594
DB	524	SREGCVTHCNFLQGEPRFVHEAGCCSCHECCLPMGTSICNCGSSGDACACAHFDPG	583
QY	595	FCVARGSGVGPDIYSTYPIKFPDEBEGACQPCPIINTCHSCVDLDDKCGPAPQASPLTS-	653

Db 584 HCVNCSCHGILG-AKGPYKYPDQONERPCHEHNTQCCNGPEIQQDCCG-AVILWSK 639  
 Qy 654 --I FNNFVSWLRVKKVSAHLEKRRQKIRKRYTRRLQETELVEPLTPSGAMPNOAQ 711  
 Db 640 PLVIAVATGLAVIIMILGSGFLWYGRRIQNKRAMRRLGEGESIEPLDPS-EKANKVL 698  
 Qy 712 KMLKETEELRKVKVLGSSAFVYVIGIWPDEENKIPAIKVLBNKSPKANKETLDEA 771  
 Db 699 ARIFKETELRKLVGSGVGVTHKGIWIPGESIKIPICIVIEDKSGRQSFQAVTHM 758  
 Qy 772 YMAVGSPVSRLLGICLTSTVQVLTQMLPGCLLDHRENGRLGSDLLNMCMOIAQ 831  
 Db 759 LAVGSLDHAHIVRLILGCGSSLGQVLTQVLPGLSLDHYKCHRETLGQLLNKVGQIAK 818  
 Qy 832 GMSYLEDVRLVHDLAARVLYKSPNHVKITPGLARLLDIDETVHADGKVPKIMMAL 891  
 Db 819 GMYLLEHSMVHRDLARVYMLKSPQVADFGVADLLPPDKOLLHSEAKTPKIMMAL 878  
 Qy 892 ESTLRRTFHOSDVMSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPQPICTIDV 951  
 Db 879 ESTHFGKTHOSDVMSYGVTVWELMTFGAEPIAGRLAEIPDLLEKGERLPQPICTIDV 938  
 Qy 952 YIMVYKMWIDSECRREFELVSEFSRNARDQRFVIONEDLGASPLDSTFYRLLED 1011  
 Db 939 YIMVYKMWIDENIRPTFKELANEFTRVARDPRYLVIKRAK-GRGTP-PAAPPSVLT 995  
 Qy 1012 DDMGDLVDAEYLVPOQGFPCDPRAGAGVHHHRSSSTSGGDLTLGLEBEE--- 1068  
 Db 996 KEL- -OEAELEPEL- - - - -DLDDLLEABEELGA 1021  
 Qy 1069 - - - - -EAPRSPLAPSEG- - - - -AGSDVFDGDLGMAKAKQISLPT 1103  
 Db 1022 TSLGSAIPLPTGLTRPRSGQLLSSSYMPMNGSSLGCAELDANVAGRRQFRRPISL 1081  
 Qy 1104 HPPSPLOKRYSEPTPLPSETGYV- - - - -APV- - - - -TC- - - - -SPOPE- - - - -YVNOQ 1143  
 Db 1082 H-FIPRGR- - - - -PASESEGHVTSSEAELOEKVYCRSRSRSPRPRDSAYHQR 1133  
 Qy 1144 DVRRPOPSPREGR- - - - -LPAARPAATLEPAKTUSP-KKNGV- - - - -KVVF 1185  
 Db 1133 HSLTPVTPPLSPGLEEDNGNYVPDTHLRGAASSREGTSSVGLSSVLTGTEEDD- - 1191  
 Qy 1186 AFGAVENEYVLPQGAAPORHP 1210  
 Db 1192 - - - - -EEYEMNRKRKRGSP-PRPP 1209  
 RESULT 11  
 EGFR\_DROME STANDARD; PRT; 1426 AA.  
 AC P04412; O61601; Q9W2G0; P81868;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)  
 DE (Gurken receptor) (torpedo protein) (Drosophila relative of ERBB).  
 GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 NC NCB1\_Taxid=7227;  
 RN (1)  
 RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).  
 RX MEDLINE=94350209; PubMed=8070664;  
 RA Clifford R., Schubach T.;  
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals  
 RT that several genetically defined classes of alleles cluster in  
 RT subdomains of the receptor protein."  
 RL Genetics 137:531-550(1994).  
 RN (2)  
 RP REVISIONS.  
 RA Clifford R., Schubach T.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE=85124611; PubMed=2982499;  
 RA Linne E., Glazer L., Segal D., Schlesinger J., Shilo B.-Z.;  
 RT "The Drosophila EGF receptor gene homolog: conservation of both  
 RT hormone binding and kinase domains".  
 RL Cell 40:599-607(1985).  
 RN [4]  
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.  
 RC STAIN-ORIGIN-R; TISSUE=Embryo;  
 RX MEDLINE=87002474; PubMed=3093080;  
 RA Schneider E.D., Segal D., Glazer L., Shilo B.-Z.;  
 RT "Alternative 5' exons and tissue-specific expression of the  
 RT Drosophila EGF receptor homolog transcripts".  
 RL Cell 46:1091-1101(1986).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=99102120; PubMed=9882502;  
 RA Leeshkin A.M., Yu S.-Y., Katz J., Baker N.E.;  
 RT "Several levels of EGF receptor signaling during photoreceptor  
 RT specification in wild-type, Ellipse, and null mutant Drosophila."  
 RL Dev. Biol. 205:129-144(1999).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I).  
 RC STAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hopkins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.Y., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.M., Benos P.V., Berman B.P., Bhandari D., Bolashkov S.,  
 RA Borkov A., Botchan M.R., Bouck J., Brinkstein P., Brotilier P.,  
 RA Butts K.C., Buzam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Idegawa C.,  
 RA Jaisl M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Mehlerson D.,  
 RA Mekulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Renkert K., Remington K., Saunders R.D.C., Scheer F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spaulding A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster".  
 RL Science 287:2185-2195(2000).  
 RN [7]  
 RP SEQUENCE OF 959-1078 FROM N.A.  
 RC STAIN=Daekwanryeong;  
 RX MEDLINE=85137938; PubMed=2983232;  
 RA Madsen S.C., Vincent W.S. III, Blodau-Wentworth D.;  
 RT "A Drosophila genomic sequence with homology to human epidermal

RT growth factor receptor";  
 RL Nature 314:178-180(1985).  
 RN [8]  
 RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=92038942; PubMed=1936959;  
 RA Raz E., Schejter E.D., Shilo B.Z.;  
 RT "interallelic complementation among DER/flb alleles: implications for  
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";  
 RL Genetics 129:191-201(1991).  
 RN [9]  
 RP REVIEW.  
 RX MEDLINE=97248481; PubMed=9094709;  
 RA Perrimon N., Perkins L.A.;  
 RT "there must be 50 ways to rule the signal: the case of the Drosophila  
 RT EGF receptor";  
 RL Cell 89:13-16(1997).  
 CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,  
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-  
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.  
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE  
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL  
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE  
 CC ESTABLISHMENT OF VENTRAL CELL FATES. MAINTENANCE OF AMNIOSESOSA  
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE  
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF  
 CC CUTICLE.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I  
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE  
 CC PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS, TYPE I (SHOWN HERE), TYPE II AND  
 CC TYPE III; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: UNICITOUSLY EXPRESSED IN EMBRYOS, IN LARVAE,  
 CC IN UTERO EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGE OF  
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST  
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW. LEVELS REMAIN HIGH  
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN  
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX  
 CC AND THORACIC AND ABDOMINAL GANGLIA.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF052754; AAC08536.1; -  
 DR EMBL; AF052753; AAC08536.1; JOINED.  
 DR EMBL; AF052754; AAC08535.1; -  
 DR EMBL; AF052752; AAC08535.1; JOINED.  
 DR EMBL; K03054; AAA51460.1; -  
 DR EMBL; K03417; AAA51460.1; -  
 DR EMBL; K03416; AAA50965.1; -  
 DR EMBL; K03418; AAA51461.1; -  
 DR EMBL; AF109077; AAD26134.1; -  
 DR EMBL; AF109078; AAD26132.1; -  
 DR EMBL; AF109082; AAD26132.1; JOINED.  
 DR EMBL; AF109078; AAD26133.1; -  
 DR EMBL; AF109084; AAD26133.1; JOINED.  
 DR EMBL; AF109079; AAD26130.1; -  
 DR EMBL; AF109081; AAD26130.1; JOINED.  
 DR EMBL; AF109079; AAD26131.1; -  
 DR EMBL; AF109083; AAD26131.1; JOINED.  
 DR EMBL; AF109080; AAD26135.1; -  
 DR EMBL; AF003454; AAF46732.1; -  
 DR EMBL; X02293; CAA26157.1; -  
 DR EMBL; X78920; CAA55523.1; -  
 DR EMBL; X72918; CAA55521.1; -

DR EMBL; X78919; CAA55522.1; -  
 DR PIR; A00640; GQFFE.  
 DR HSPD; P11362; 1EGK.  
 DR FlyBase; FBgn0003731; Egfr.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_Pkinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR PRODOM; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SM00261; Fur; 7.  
 DR SMART; SM00219; TyrKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 DR PROSITE; PS0011; PROTEIN KINASE DOM; 1.  
 DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;  
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;  
 FT SIGNAL 1 30  
 FT CHAIN 31 1426  
 FT DOMAIN 31 869  
 FT TRANSMEM 869 889  
 FT DOMAIN 938 1426  
 FT DOMAIN 938 1198  
 FT NP\_BIND 944 952  
 FT BINDING 971 971  
 FT ACT\_SITE 1063 1063  
 FT MOD\_RES 902 902  
 FT PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 Query Match 28.3%; Score 1936; DB 1; Length 1426;  
 Best Local Similarity 32.7%; Pred. No. 5,8e-95;  
 Matches 467; Conservative 183; Mismatches 432; Indels 346; Gaps 43;  
 24 QVCGTGMKRLPASPETHLDMRLHYGGQGVQVQGNLELYLP-NASLSFLDIOGVG 82  
 100 KICGTSKRSVSPENKHHYRNLDRITNCTYDGNLKLWLPENLDSFLNIREVTG 159  
 83 YVLIHQVQVQVPLQSLRIVRGTOF-----EDNYALAVLDNGDPLNNTTPVYGASDGL 137  
 160 YLISHVDVKKVPEPKQIIRGRTLFSLVEEEKYALFV-----TYSKM 203  
 138 RELQSLTELKKGVLQKNPQLCYDITLMDIFKKNQALTLIDTRSRACHCSP 197  
 204 YLLEIPDLRLVNGQVGFHNHNYMLCHRRTIQWSEIYVNGDAYYNDFTAPRECPRCH 263  
 198 MCKSRRCWGESSEDCQSLRTVACAGCA--RCKGPLPTDCGHECAGCTGPKHSDCLAC 255  
 264 SCTHG-CWGBSPKNCQFSLKLTGSPQAGRCYCPKRECHLFCAGCTGTPQKDCIAC 322  
 256 LHFNSGICELHCPALTYNTDTFESMPNEGRYTFGASCVTACPVNYISTDVGSCITLVC 315  
 323 KNFDEAVSKCECPMKRYPTTYLTETNEBGKAYVATCVKECP-GHLRLDNGACVRSK 381  
 316 PLNHQVTAEDGTGRCKCKPKPARVYGVGQVYKANSFPGITEL-----EFAGCKK 369  
 382 PDDMDKGE-----CVPCNGPCKTC-----PGTVLHAAGNISFRNCTV 422  
 370 IFGSIAPLPESFDS--DPAANTA-----FLOPEQLQVFTLEITGYLYISAMPDLPD 421  
 423 IDGNIRILDQTFSSFCQVYVANYTWGPRTYPLDPERRREVFTVKEITGYLTNTEGHPQFRN 482  
 422 LSVQNIQVIRGRILNHGAY-SLTQQLGISMGLRLRLRELGLALIHNNTHLCFVHTV 480  
 483 LSYFRNLFTIHGRQLMSMPALAIYSSLSYLSLEMRLLKQISSGVVYIQHNRDLQVYSNI 542  
 481 PMDQLFRNPQALHTANRPEDEVCGEGLACHQLCARHGWCWGPPTQVACQSPFLRQEC 540  
 543 RMPAIQKEPEBKVWVNNLADLCENKGTICSDQCNNDGCGWAGTQCLCKKPNFRGTC 602  
 541 VEECRVLIQGLPREYVNAHRLCPHPECQPNQNGSVTCGPEADQCACAAHYKDPPEVCARC 600

```

Db      603 IADCGYISNAK--FONRTCKICHPEER-----TCNGAGADHCQCEVHVDGQHCVSEC 654
Qy      601 P-----SGVK-----PDL 608
Db      655 PKNKYNDRGVRECHATCDGCTGPKDTIGIGACTGCTNIAINNDATVKSCLLKDRCPD- 713
Qy      609 SYMPIKFE--PDEEGACQ-----CPI-----NCTH-----632
Db      714 GY--FWEYVHPQEGSLKPLAGAVCRKCHPLCELCNTNNGYHQCSCKTHYKRRQECT 771
Qy      633 -----SC-----VDLDKG-----641
Db      772 ECPADHYTDEEQRCEQHPHCEGCTPGADCKSCRNFPLPANTGTGVNSTMENCT 831
Qy      642 -CPAQR-----ASPLS--TFNN-----FTYSFWLRVPKVSASHLEK-- 676
Db      832 KCPLEMHVNYQYTAIGPYCAASPPRSSKITANLDVNMIFILIGAVLVPTICILCVVYTI 891
Qy      677 -RRQOKIRKYT--MRRLQETELVEPLTPSGAMPNOAMRILKETELRKVKVLSGSAFGT 733
Db      892 CRQKQAKKETVMTALSCGDESEPLRPSNIGANCKLRIVDADALRKGVLGMAFAFR 951
Qy      734 VYKGIWIPDGENVKIPVAILKRENTSPKANKIIDEAYVMAGVSPVSRLLGICLTST 793
Db      952 VYKGVWVPEGENVKIPVAILKELKSTGAESEFEFLREAYIMASEEHVNLKLLAVCMSQ 1011
Qy      794 VQVLTQMPYGCCLDHYRENRGSLGSODLLNMCQIAKMSYIEDVRLVHRDLAANVLY 853
Db      1012 KMLITQMLPGCLLDVVRNRRKDISKALNMSTQIAKMSYLEERLVRDLAANVLY 1071
Qy      854 KSPNHVKITDFGLARLLIDETEVHADGKVPKIMMALESILRRRPTHOSDWSYGVTVW 913
Db      1072 QTSFLVKITDFGLAKLLSDSNENYKAGGKMPKIMLALCINRNVTSISDVMARGVTTM 1131
Qy      914 ELMTFAKRPYDGIIPAEIPDLLEKGRLLQPPICITDVYMWKCMKIDSECRPREELV 973
Db      1132 ELTFQGRPHENIPADIPDLLEVGKLEKQPELCSIDICTLLSCHMLDAAMRPFKQLT 1191
Qy      974 SEFSRMADPQRFVVIQNEIDL--PASPLDSTFYRSLDEDD--DWGDLVDAEVLVPOQ 1028
Db      1192 TYFAERARDPGRYLAIPGDKFTRLP-----YTSQDCKDLIRKLAFTDSSEAIAPD 1244
Qy      1029 GFPCPDPAQAGQVHRRHRSSTRSGGDLTLGLEPSEEAR-----RSLPASEGAG 1082
Db      1245 DYLOPKAAPGS-----HRTDCT-----DEMPKLNRYCKDPSNKNSSGTG 1283
Qy      1083 SDVFDG--DLMGAAKGLQSLPTHDPSPLOKRYSENDPTVPLPSEFDGVYAPLTGSPQBY 1135
Db      1284 DDERDSSAREVGVGNLR-----LDLPVDEDDYLMPTCOFGPNN 1321
Qy      1140 VNQPDVVRPOPSPREGPLPARAPAGATLERAKTLPQKGVGVDFAFGAVENPEYL-- 1197
Db      1322 NNMMN-----NPNQNNMAAVGVAAGM-----DLGVVSDNDEYLLN 1360
Qy      1198 --TPQGAAPQH-----PPAFSP-APDNLYYWD 1224
Db      1361 AQLTGVGSPITPTQITIGIPVWGPGTMEVYKPMPSSEPTSSDHEYND 1408

```

```

RESULT 12
ERRB ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein etdb (EC 2.7.1.112).
GN V-ERBB
OS Avian leukosis virus
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_TaxID=11864;
RN [1]

```

```

RP      1012 KMLITQMLPGCLLDVVRNRRKDISKALNMSTQIAKMSYLEERLVRDLAANVLY 1071
RX      MEDLINE=85228222; PubMed=2988784;
RA      Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA      Crittenden L.B., Raines M.A., Kung H.-J.;
RT      "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT      processing and promoter insertion result in expression of an
RT      amino-truncated EGF receptor."
RL      Cell 41:719-726 (1985).
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC      IN CHICKS WHICH HAVE BEEN INFECTED WITH THE AVIAN LEUKOSIS VIRUS
CC      AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC      THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC      C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC      -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC      PROTEIN.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib.ch).
CC      EMBL, M10066; AAA8763.1; ALT_INIT.
CC      PIR, A0643; TVCHLV.
CC      PIR, B0643; TVFVLV.
CC      HSPD, P1362; 1FGK.
CC      InterPro, IPR000719; Euk_Dkinase.
CC      InterPro, IPR001245; Tyr_kinase.
CC      Pfam, PF00069; Pkinase; 1.
CC      PRINTS, PR00109; TYRKINASE.
CC      Prodom, PD000001; Euk_Dkinase; 1.
CC      SMART, SMO0219; Tyrc; 1.
CC      PROSITE, PS00107; PROTEIN KINASE ATP. 1.
CC      PROSITE, PS00109; PROTEIN_KINASE_TYR. 1.
CC      PROSITE, PS00011; PROTEIN_KINASE_DOM. 1.
CC      Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC      Glycoprotein; Phosphorylation.
CC      FT DOMAIN 132 399
CC      FT NP BIND 138 146 ATP (BY SIMILARITY).
CC      FT BINDING 165 165 ATP (BY SIMILARITY).
CC      FT ACT SITE 257 257 BY SIMILARITY.
CC      SEQUENCE 634 AA; 70891 MW; E705B33A0BE01FCC CRC64;
Qy      Query Match 24.9%; Score 1704.5; DB 1; Length 634;
Qy      Best Local Similarity 51.6%; Pred. No. 4e-83;
Qy      Matches 364; Conservative 72; Mismatches 149; Indels 121; Gaps 15;
Qy      587 CAHYKDPFCVYARCPGKVDLSYMPIKFPDEEGACQCPINCTHSCVLDLDKCPAQ 646
Qy      3 CAHFIDGPHCVKACPAVLGENDTL-VWKYADANAVALQCHPNTCRCKGPGLEGCP--- 58
Db      647 RASPLTSFNNFTVSFVLRYPKVASLLEKRRQKIRKYMRLQETELVEPLTPSGAM 706
Db      59 NGSKTPIAAGVVGCLLVVVGIGLYLRRRIYKRTLRLLQERELVEPLTPSGEA 118
Qy      707 PNOQMRILKETELRKVKVLSGAFGTVYKGIWIPDGENVKIPVAILKRENTSPKANK 766
Db      119 PNOHRLILKETELRKVKVLSGAFGTVYKGIWIPDGENVKIPVAILKRENTSPKANK 178
Qy      767 ILDEAYVMAGVSPVSRLLGICLTSTVQVLTQMPYGCCLDHYRENRGSLGSODLLNMC 826
Db      179 ILDEAYVMASVDNPHVCRLLGICLTSTVQVLTQMPYGCCLDHYREKDNIGSQYLINMC 238
Qy      827 MOIAKMSYLEDVRLVHRDLAARVLYKSNHVKITDFGLARLLIDETEVHADGKVP 886
Db      239 VOIAKGNVYLEERLVRDLAANVLYKTOHKITDFGLAKLGADEKYYHAEGGKVP 298
Qy      887 KMALESILRRRPTHOSDWSYGVTVWELMTFGAKPYDGIIPAEIPDLLEKGRLLQPP 946

```

```

Db 299 KMALESILHRIYTHQSDVMSYGVYVWELMTFGSKPYDGIASEISVLEKGERLPQPI 358
Qy 947 CTIDVYIMVCMWIDSECRPRELVSSESRMARBPORVVIQ-NEDLGASPLDSTFY 1005
Db 359 CTIDVYIMVCMWIDSECRPRELVSSESRMARBPORVVIQ-GERMHLPSPTDSKY 418
Qy 1006 RSLLEDDMDGLVDAEELVLPQGFPCPDPAAGGVHHRHSSSTRSGGDLTLGLEP 1065
Db 419 RLMEEEDMEDIVDADELVPHQGF-----NSPST----- 449
Qy 1066 SEEBAPRSPL-----APSEGASDVDFDGLGMAAGLQSLPHDPSPLQRYSEDPVPL 1120
Db 450 -----SRPLSLSSLSATSNNSATNCID-----RNGQGHPRVEDSFVQRYSSDPTGNF 496
Qy 1121 PSET--DGYVAPLTCSPQPEYVNOPDVPRPSPREGPLPAARPAAGTLERAKTLSPGKN 1178
Db 497 LEESIDGFL-----PAPRYVNO--LMPKKPS-----TAMVON 527
Qy 1179 GVYKDVAF-----AFGAIVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYM 1223
Db 528 QIYNNISLTAISKLPMDSRYSNGSHSTAVDNPEYL-----NTNOSPPLAKTVFESSPYW 579
Qy 1224 DQ-----DPEE-----RGAPSTFKGTPTAENPEYLGIDVP 1254
Db 580 IQSGNHQIINLDNPDYQODFLPNETTKFVGLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER
ID ERBB_AVIER STANDARD; PRT; 604 AA.
AC P0053;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain BS4).
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxId=79685;
RN 1)
RN SEQUENCE FROM N.A.
RC STRAIN=H;
RX MEDLINE=4026539; Pubmed=6333229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbb gene of avian erythroblastosis virus is a member of the src
RT gene family.";
RL Cell 35:71-78(1983).
RN 2)
RN SEQUENCE OF 1-152 FROM N.A.
RX MEDLINE=4223957; Pubmed=6328658;
RA Debure B., Henry C., Benaisse M., Biserte G., Claverie J.-M.,
RA Seale S., Martin P., Stehelin D.;
RT "Sequencing the erbb gene of avian erythroblastosis virus reveals a
RT new type of oncogene.";
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

```

```

DR EMBL: K02006; AAA42394.1; ALT_INIT.
DR EMBL: K01216; AAA42400.1; -.
DR PIR: A00644; TVPK.
DR HSPF: F11362; TVPK.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PF00069; kinase; 1.
DR ProDom: PD000001; Euk_kinase; 1.
DR SMART: SM00219; TyrKc_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 S->W (IN REF. 2).
FT CONFLICT 140 140 R->F (IN REF. 2).
FT CONFLICT 146 146 I->V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76BECDD06745D609 CRC64;

```

```

Query Match 24.3%; Score 1658; DB 1; Length 604;
Best Local Similarity 51.4%; Pred. No. 11e-80;
Matches 354; Conservative 69; Mismatches 142; Indels 124; Gaps 14;

```

```

Qy 587 CAHYKDPFCVAPRCSPGVPLDSYMIKWPDESGACQPCPINCSTHSCVDLDDGCCPAE 646
Db 3 CAHFDGPHCVKACPAAGVLTGENDTL-VKIVADANAACVCLCPNCTRGCKGGLGCP-- 58
Qy 647 RASPLTSIFNNFTVSFWLVKPVKASASLEKRRQCKIRKRYTRRLLOETELVEPTLPSGAM 706
Db 59 NGSKTPSIAGWVGGLCLLVVWGIGLYLRRRIIVRRTLRLLQGERELVEPTLPSGEA 118
Qy 707 PNOGMRILKTELKRVVLGSGAFGYTKYKIMIPDGENVKIPVAIKYLRNTSPKANKE 766
Db 119 PNOALRLKTEFEKRVVLGSGAFGITYKLMIPGEKVKI PAIKELRATSPKANKE 178
Qy 767 ILDEAYVWAGSPVYSRLIGICTSTVQVLTQMLPQGLLDHYRENGRLGSDLLNMC 826
Db 179 ILDEAYVWASVDNHHVCHLIGICTSTVQVLTQMLPQGLLDYRREHNDNGSYLLNMC 238
Qy 827 MOIAKMSYLDVRLRHDLAARVLYKSPNHVKITDPGLARLLDIDETVHADGKVP 886
Db 239 VOIAKMYVLEERLVRHDLAARVLYTPQHVKITDPGLAKLLGADEKEVHAEGKVP 298
Qy 887 KMALESILHRRFTHOSPDVMSYGVYVWELMTFGSKPYDGIASEISVLEKGERLPQPI 946
Db 299 KMALESILHRIYTHQSDVMSYGVYVWELMTFGSKPYDGIASEISVLEKGERLPQPI 358
Qy 947 CTIDVYIMVCMWIDSECRPRELVSSESRMARBPORVVIQ-NEDLGASPLDSTFY 1005
Db 359 CTIDVYIMVCMWIDSECRPRELVSSESRMARBPORVVIQ-GERMHLPSPTDSKY 418
Qy 1006 RSLLEDDMDGLVDAEELVLPQGFPCPDPAAGGVHHRHSSSTRSGGDLTLGLEP 1065
Db 419 RLMEEEDMEDIVDADELVPHQGF-----NSPST----- 449
Qy 1066 SEEBAPRSPL-----APSEGASDVDFDGLGMAAGLQSLPHDPSPLQRYSEDPVPL 1120
Db 450 -----SRPLSLSSLSATSNNSATNCID-----RNGQGHPRVEDSFVQRYSSDPTGNF 496
Qy 1121 PSET--DGYVAPLTCSPQPEYVNOPDVPRPSPREGPLPAARPAAGTLERAKTLSPGKN 1178
Db 497 LEESIDGFL-----PAPRYVNO--LMPKKPS--TAMVON 524
Qy 1179 GVYKDVAF-----AFGAIVENPEYLTPOGGAAPQHPHPPAFSPAFDNLVYM 1223
Db 528 QIYNNISLTAISKLPMDSRYSNGSHSTAVDNPEYL-----NTNOSPPLAKTVFESS 575
Qy 1220 LYWDQDPPERGAPPSFTFKGTPTAENPEY 1248

```

Db 576 SPYWISQGNHQ-----INLDNDY 594

RESULT 14

ERBB\_AVIEW

ID ERBB\_AVIEW STANDARD; PRT; 540 AA.

AC P11273;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).

GN V-ERBB.

OS Avian erythroblastosis virus (strain ts167).

OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.

ON NCBI\_TaxId=103898;

RP SEQUENCE FROM N.A.

RA MEDLINE=87064458; PubMed=2978364;

RA Choi O.R., Trainor C., Graf T., Beng H., Engel J.D.,

RT "A single amino acid substitution in v-erbB confers a thermostable

RT phenotype to ts167 avian erythroblastosis virus-transformed erythroid

RT cells."

RT Mol. Cell. Biol. 6:1751-1759(1986).

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC tyrosine phosphate.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M13179; AAA2401.1; -

DR PIR: A25231; TYFVB.

DR HSSP: P1362; 1FGK.

DR InterPro: IPR000719; Euk\_pkinase.

DR InterPro: IPR001245; Tyr\_pkinase.

DR Pfam: PF00069; pkinase; 1.

DR ProDom: PD000001; Euk\_pkinase; 1.

DR SMART: SMO0219; Tyrc; 1.

DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.

DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.

DR Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;

DR Glycoprotein; Phosphorylation.

FT DOMAIN 132 389

FT NP BIND 138 146

FT BINDING 165 165

FT ACT SITE 257 257

FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).

FT SEQUENCE 540 AA; 60412 MW; 5B53297AA06865D CRC64;

Query Match 23.2%; Score 1585; DB 1; Length 540;

Best Local Similarity 54.2%; Pred. No. 6,8e-77;

Matches 335; Conservative 62; Mismatches 133; Indels 88; Gaps 12;

587 CAHYDPFCVAVRCGVPDLVYPIKPFDEGACQPCPINCNCSCVDLDDGCPAPQ 646

Db 3 CAHFDGPHCVACAGVAGLGDNL-VKRYADANVCLCPNCRGKGGLESCP--- 58

QY 647 RASPLTSTFNNFTVSFMRVPKVSASHLEKRRQKIRYRYMRLLQETELVEPTPSGAM 706

Db 59 NGSKTPSIAGVVGGLLVVVGIGLYLRHRHVRKTRRLRLLQERELVEPTPSGCA 118

QY 707 PNOAGRIKETELRKVKVLSGARGTYKGIWIPDGNVYIPVATYLRNTSPKANK 766

Db 119 PNOAHLRIKETEFKVKVVLGSGAGTYKGMIPGSKVITPAIKELREATSPKANK 178

QY 767 ILDEAYVAGVGSPPVSRLLGLCTSTVQLVTLQMPYGCCLLDHVENNGRLGSDLLNMC 826

Db 179 ILDEAYVNASVDNHRVCRLLGLCTSTVQLTQMPYGCCLLDYRHKNDIGSOYLNMC 238

QY 827 MOIAKMSYLEDVALVHRDLAARNVLYKSPNHWKITDGLARLLDIDETEHADGKVP 886

Db 239 VOIAKMNVLBERHWHRLDARNVLYKTPQHVKITDGLAKQGADEKEYHAGKVP 298

QY 887 KMMLESLRRRFLTHQSDVWSYGVTELMFTFGAKPYDGIIPARITPDLLEGERLPQPI 946

Db 299 KMMLESLHRIYHQSDVWSYGVTELMFTFGSKPYDGIIPASISVLEGERLPQPI 358

QY 947 CTIVVMIMVCMWIDSECRFPRELSEFSRMRDPPRFVIO-NEDLGASPLDSFTFY 1005

Db 359 CTIVVMIMVCMWSDADSRKFFELIAFESKWARDPPRYIVIGDERMLPSTDSFTFY 418

QY 1006 RSLLEDMDGLVDAEYLVPQGFPCPDPAAGVNHHRSSSTRSGGDLTLGLEP 1065

Db 419 RLMESEDMEDIVDAEYLVPHQGF-----NSPST----- 449

QY 1066 SEEDAPRSP-----APSEGAGVDFPDGDMGAQGLQSLPTHPDPSPLQYSDDPTVPL 1120

Db 450 -----SRTPLSLSLSTSNNSATNCIRNGS-----H----- 476

QY 1121 PSETDGYVAPLTCSPQEVYVQPDVRFQPSFREGPPAPAPAGAT-LEPAKTUSPGNG 1179

Db 477 PVREDGFL-----PADEVYVQ--LMPKXSTAMVQVQIVYISLTSKLPIDSRVQN- 527

QY 1180 VKQVFAFGAVENPEYL 1197

Db 528 -----SHSTAVDNPEYL 539

RESULT 15

EGFR\_CHICK

ID EGFR\_CHICK STANDARD; PRT; 703 AA.

AC P13387;

DT 01-JAN-1990 (Rel. 13, Created)

DT 01-JAN-1990 (Rel. 13, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER

DE Fragment).

GN EGFR.

OS Gallus gallus (chicken).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;

OC Gallus.

OX NCBI\_TaxId=9931;

FN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=88261272; PubMed=3260329;

RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,

RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.,

RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,

RT expression in mouse cells, and differential binding of EGF and

RT transforming growth factor alpha."

RT Mol. Cell. Biol. 8:1970-1978(1988).

CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF.

CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP130 AND

CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC tyrosine phosphate.

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to

CC dimerization, internalization of the EGF-receptor complex,

CC induction of the tyrosine kinase activity, stimulation of cell DNA

CC synthesis, and cell proliferation.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@sb.sib.ch).

CC -----  
 DR EMBL: M20386; AAA48760.1; -  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR SMART: SM00261; Fu; 4.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; PARTIAL.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; PARTIAL.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; PARTIAL.  
 DR Transmembrane; Glycoprotein; Receptor; Signal; transferase;  
 KM Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
 FT SIGNAL 1 30  
 FT CHAIN 1 >703  
 FT DOMAIN 31 654  
 FT TRANSMEM 655 667  
 FT DOMAIN 668 >703  
 FT DISULFID 197 206  
 FT DISULFID 201 214  
 FT DISULFID 222 230  
 FT DISULFID 226 238  
 FT DISULFID 239 247  
 FT DISULFID 243 255  
 FT DISULFID 258 267  
 FT DISULFID 271 298  
 FT DISULFID 302 314  
 FT DISULFID 318 333  
 FT DISULFID 336 340  
 FT DISULFID 513 522  
 FT DISULFID 517 530  
 FT DISULFID 533 542  
 FT DISULFID 546 561  
 FT DISULFID 565 581  
 FT DISULFID 589 592  
 FT DISULFID 592 601  
 FT DISULFID 605 627  
 FT DISULFID 630 638  
 FT DISULFID 634 646  
 FT CARBOHYD 134 134  
 FT CARBOHYD 190 190  
 FT CARBOHYD 200 200  
 FT CARBOHYD 200 200  
 FT CARBOHYD 359 359  
 FT CARBOHYD 368 368  
 FT CARBOHYD 420 420  
 FT CARBOHYD 573 573  
 FT CARBOHYD 578 578  
 FT CARBOHYD 613 613  
 FT CARBOHYD 633 633  
 FT CARBOHYD 648 648  
 FT NON\_TER 703  
 SQ SEQUENCE 703 AA; 77427 MW; AAFP2DE11B735A690 CRC64;

Query Match 22.7%; Score 1550; DB 1; Length 703;  
 Best Local Similarity 43.8%; Pred. No. 6,4e-75;  
 Matches 310; Conservative 104; Mismatches 265; Indels 28; Gaps 12;

QY 8 RRGILLALLPPGAA-----STQVCTGDMKRLPASPEETHLDMRLHYQGCVVQGNLE 61  
 DB 13 RGAVALVLLLLGVALCSAVEKVCQGTNNKLTQLGHVEDHFTSLQPMYNNCEVLSNLE 72  
 QY 62 LTLVLPVNASLSFLQDIEVQGVYLAHQVQVPLQRLIRVSTQTFEDRYALAVDNGD 121  
 DB 73 ILYVENHRDLTLFKTIDVAGVYLIALNMWDVIFLENQIIRGNVLYDNSFALAVLSNTH 132  
 QY 122 PLNNTPVVTGASPGGLRELOKSLTEILKGVLIQRNPOLCYQDTILMKDIFHKNNQIAL 181  
 DB 133 -NNKTO-----GLRELPMKRLSELINGVXISNNPKLCNMDTVLWMDIIDTSRK-PL 182  
 QY 182 TLID-TVRSRACHPCSPMCKSGRCWGSSEDCSLTRTVAGGCA-RCKGPLPTDCHEO 239

DB 183 TLIDPASNLSGCPKPCPNCTEDHGWGAGEQNCQILTKVICAQOCSGRCRGHPSPDCCHNQ 242  
 QY 240 CAAGCTGPKHSDCLACLHFNHSGICEHLPLVLYNTDTEFSMNPREGRYTFGASCYTAC 299  
 DB 243 CAAGCTGPRESDCLACRKFRRDATTCTGCPPLVLYNPTTYMDVNPBEKISFGATVREC 302  
 QY 300 PNYLSTDVSGCTLVCPPLHNOEYVAEDGTORCEKSKPCAVCYGLAQVYIKANSKFI 359  
 DB 303 PHNYVTDHGSCVRSCTNDTYE-EENGVRCCKKCDGLSKVNCNGIGIGELKGLIS-INA 360  
 QY 360 TELE-FACCKKIFGSLAFLESPFDGDPASNTAPLOPELOVPELTLEETGLYISAWPDS 418  
 DB 361 TIDSEFXKCTKINGDVSTLPVAFUGDAPTKTLPDPKLDVFRVYKEISGFLLLQAMPDN 420  
 QY 419 LPDLSVFNLIQVIRGRIIHNAGVASTLQGLGISWLSRLSELGSLALIHNNHTLCFVH 478  
 DB 421 ATDLYAFENLEIRRTQHOQYSLAVNVLKQISGLRSLKEISGDLAIKKNKRLCTAD 480  
 QY 479 TVPMDOLEFRNPQALLHTANRPEDECVGEGLAGHQLCARGHGWGPGPTQCVNCSQFLRGQ 538  
 DB 481 TMANRSLPATOSQKTKIIONENKNDCTADRHVCDPLCSDVGCWGGPPEHCFSRFRBQK 540  
 QY 539 EGVESCVTLQGLPREYVNAARHCLPCHPECPONG---SVTQGPPEADQCVCAHYKDPF 595  
 DB 541 ECVKQCNLTQGBPREFERDSKCLPCHSECLVQNSTAYNTTSGRPPDHCKCAHFIQDPH 600  
 QY 596 CVARCPGSKVDLSYMPIWKPEDEGACQPCPINCTHSCVDLDKGCFAEQRASPLTSIF 655  
 DB 601 CVKACPAGVLENDTL-VMKYADANAVALQLHPNCTRCCKPGLEGCP---NGSKTPSIA 656  
 QY 656 NNFTVSPWLRPKVASASHLEKRRQCKIKYTMKRLDQTELYEPLTP 702  
 DB 657 AGVGGLLCLVVGVLGIGLYLRRIYVKKRLRLRLORRELYEPLTP 703

Search completed: July 22, 2003, 09:18:05  
 Job time : 21.642 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.9246 Seconds

(without alignments)  
5522.503 Million cell updates/sec

Title: SEQ4-632-652-14

Perfect score: 6834  
Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTNENPEYGLDVPV 1255

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_RODENT:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6125	89.6	1259	6	018735 canis fami
2	3107	45.5	1209	1	09GX70
3	3078	45.0	1210	11	09EP98
4	2685	39.3	1165	13	09YH40
5	2646.5	38.7	1137	13	Q9W674
6	2260	33.1	1328	13	P79754
7	1988.5	29.1	1433	5	09B1H9
8	1871	27.4	419	4	09UK79
9	1739	25.4	367	11	08R2X1
10	1697.5	24.8	412	4	08MYV0
11	1675	24.5	729	15	086712
12	1673	24.5	567	15	086714
13	1608.5	23.5	962	15	064895
14	1600	23.4	545	15	085468
15	1506.5	22.0	655	11	09WVFS
16	1490.5	21.8	643	11	09ERV6

17	1247	18.2	1193	5	Q9Y1X8	Q9Y1X8 ephydrat f
18	1181.5	17.3	1368	5	Q23821	Q23821 caenorhabdi
19	1150	16.8	1717	5	Q26566	Q26566 schistosoma
20	1126	16.5	527	13	Q90836	Q90836 gallus gall
21	1001.5	14.7	478	11	Q9E8E0	Q9E8E0 rattus norv
22	942.5	13.8	599	13	Q9P8H2	Q9P8H2 gallus gall
23	906	13.3	165	4	Q14256	Q14256 homo sapien
24	887	13.0	176	11	Q923V5	Q923V5 rattus norv
25	806.5	11.8	346	13	P11776	P11776 xiphophorus
26	778	11.4	435	5	Q8SZW1	Q8SZW1 drosophila
27	754.5	11.0	311	13	Q99162	Q99162 xiphophorus
28	739.5	10.8	1362	13	Q9PV24	Q9PV24 xenopus lae
29	734	10.7	331	4	Q9BUD7	Q9BUD7 homo sapien
30	723	10.6	149	6	Q9BG66	Q9BG66 oryctolagus
31	715	10.5	1671	5	Q9NJV5	Q9NJV5 biomphalaria
32	688	10.1	1418	13	Q93457	Q93457 scophthalmu
33	675.5	9.9	1368	13	Q8UW85	Q8UW85 paralicthy
34	673.5	9.9	1369	13	Q8UW86	Q8UW86 paralicthy
35	657.5	9.6	1472	5	Q9U5A8	Q9U5A8 bombyx mori
36	657	9.6	1358	13	Q73798	Q73798 xenopus lae
37	652	9.5	1412	13	Q8UW84	Q8UW84 paralicthy
38	635.5	9.3	1418	13	Q8UW83	Q8UW83 paralicthy
39	631	9.2	1245	13	Q9YCH8	Q9YCH8 scophthalmu
40	629	9.2	1371	11	Q9OYH4	Q9OYH4 rattus sp.
41	621.5	9.1	2144	5	Q9VD54	Q9VD54 drosophila
42	589.5	8.6	1091	4	Q9UMQ4	Q9UMQ4 homo sapien
43	587.5	8.6	1036	4	Q07912	Q07912 homo sapien
44	586.5	8.6	1055	11	Q54967	Q54967 mus musculu
45	580.5	8.5	1072	4	Q9BTD0	Q9BTD0 homo sapien

## ALIGNMENTS

RESULT 1  
ID 018735 PRELIMINARY; PRT; 1259 AA.  
AC 018735;  
DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
DE Erib-2.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Yokota H.;  
RT "CDNA cloning of erib-2 from canine mammary gland."  
RL Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AB008451; BAA23127.1; -  
DR HSSP; P11362; 1FGK.  
DR InterPro; IPR002048; EF-hand.  
DR InterPro; IPR000494; EGFR\_L\_domain.  
DR InterPro; IPR000719; Euk\_Pkinase.  
DR InterPro; IPR002174; Furin-like.  
DR InterPro; IPR001245; Tyr\_Pkinase.  
DR InterPro; IPR004019; YLP\_motif.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; Pkinase; 1.  
DR Pfam; PF01030; Recep\_L\_domain; 2.  
DR Pfam; PF02757; YLP; 2.  
DR ProDom; PD000001; Euk\_Pkinase; 1.  
DR SMART; SM00261; FU; 3.  
DR SMART; SM00219; TYKc; 1.  
DR PROSITE; PS00018; EF\_HAND; UNKNOWN; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00101; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
KW ATP-binding; Transferase; Tyrosine-protein kinase.  
SQ SEQUENCE 1259 AA; 137969 MW; E37364D4C4ACD46 CRC64;

Query Match 89.6%; Score 6125; DB 6; Length 1259;  
 Best Local Similarity 89.8%; Pred. No. 0;  
 Matches 1132; Conservative 40; Mismatches 92; Indels 6; Gaps 2;

QY 1 MELALCRMGLLALIPPAASTOVCTGTDMKRLPASPETHLDMRLHYGCCQVYQNL 60  
 DB 1 MELALCRMGLLALIPPAASTOVCTGTDMKRLPASPETHLDMRLHYGCCQVYQNL 60  
 QY 61 ELTYLPNTASISLFDIOIEVQGVYLAHNOVQVPLRQRLIRYRQQLPFDNVALVLDNG 120  
 DB 61 ELTYLPNTASISLFDIOIEVQGVYLAHNOVQVPLRQRLIRYRQQLPFDNVALVLDNG 120  
 QY 121 DPLNNTPVYTGASPGGARELQRLSLTEILKGVYLGQNPQLCYDTIWKDI FAKNQLA 180  
 DB 121 DPLNNTPVYTGASPGGARELQRLSLTEILKGVYLGQNPQLCYDTIWKDI FAKNQLA 180  
 QY 121 DPLNNTPVYTGASPGGARELQRLSLTEILKGVYLGQNPQLCYDTIWKDI FAKNQLA 180  
 DB 121 DPLNNTPVYTGASPGGARELQRLSLTEILKGVYLGQNPQLCYDTIWKDI FAKNQLA 180  
 QY 181 LTLITNRSRACHPGSPMCKSGRCSGSESDCQSLTRTYCAGGACRCKPLTDCHEQC 240  
 DB 181 LTLITNRSRACHPGSPMCKSGRCSGSESDCQSLTRTYCAGGACRCKPLTDCHEQC 240  
 QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDPESPNDEGRYTFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDPESPNDEGRYTFGASCVTACP 300  
 QY 301 YNLSTDVSGCTLVCEFLHNOVYAEQTCRCEKSPCARVCYGLMOYIKANSKFIGIT 360  
 DB 301 YNLSTDVSGCTLVCEFLHNOVYAEQTCRCEKSPCARVCYGLMOYIKANSKFIGIT 360  
 QY 361 ELEFACCKIFGSLAFLEPFDGDPASNTAPLOPELOQVFEITLITGLYISAMPDILP 420  
 DB 361 ELEFACCKIFGSLAFLEPFDGDPASNTAPLOPELOQVFEITLITGLYISAMPDILP 420  
 QY 421 DLSVFONLOYIRGRIIHNQAYSLTLOGGISWLGRLSLELGSGLALHNTLCPVHTV 480  
 DB 421 DLSVFONLOYIRGRIIHNQAYSLTLOGGISWLGRLSLELGSGLALHNTLCPVHTV 480  
 QY 481 PNDOLFNRPHOALHTANRPEDECEVGLACHQLCRGHCWGMPGPOCNGSOFIRGOC 540  
 DB 481 PNDOLFNRPHOALHTANRPEDECEVGLACHQLCRGHCWGMPGPOCNGSOFIRGOC 540  
 QY 541 VEECRVLOGLPREYVNAHCLPCHRECPQNSVTCFGEADQVCAAHYKDPFCVARC 600  
 DB 541 VEECRVLOGLPREYVNAHCLPCHRECPQNSVTCFGEADQVCAAHYKDPFCVARC 600  
 QY 601 PSQVPRDLSPMPKWKADDEGTQPCPINCTHSCADLDEKGPAPBRASPVTISIAAVG 659  
 DB 601 PSQVPRDLSPMPKWKADDEGTQPCPINCTHSCADLDEKGPAPBRASPVTISIAAVG 659  
 QY 661 SEWLRVPKVSASHLEKROOKIRKYMRLLOETELVEPLTESGAMPNQAQRILKETEL 720  
 DB 661 SEWLRVPKVSASHLEKROOKIRKYMRLLOETELVEPLTESGAMPNQAQRILKETEL 720  
 QY 721 RKXKVLGSGAFGVYGIWI PDENKIVAIKVLNENSPRANKIIDEAVYMAVGSP 779  
 DB 721 RKXKVLGSGAFGVYGIWI PDENKIVAIKVLNENSPRANKIIDEAVYMAVGSP 779  
 QY 781 YVSRLLGICLTSTVQVLTQMLPFGCLDHYRENRGLSGODLLNMCQJAKGMSYLEVR 840  
 DB 781 YVSRLLGICLTSTVQVLTQMLPFGCLDHYRENRGLSGODLLNMCQJAKGMSYLEVR 840  
 QY 841 LVHRDLAARNVLYKSNHVKITDFGLARLLDIDETRYADGGKVPKMALESILRRFT 900  
 DB 841 LVHRDLAARNVLYKSNHVKITDFGLARLLDIDETRYADGGKVPKMALESILRRFT 900  
 QY 901 HOSDVMYSGVYVWELTFGAKPYDGI PAEIPDLLEKGRILQOPRTCTIDVYMIWKKCM 960  
 DB 901 HOSDVMYSGVYVWELTFGAKPYDGI PAEIPDLLEKGRILQOPRTCTIDVYMIWKKCM 960  
 QY 961 IJSECRPRRELVSFSRNARDPQRFVJQNEDLGPASLDSTFYRSLDEDDMDGLVDA 1020  
 DB 961 IJSECRPRRELVSFSRNARDPQRFVJQNEDLGPASLDSTFYRSLDEDDMDGLVDA 1020  
 QY 1021 EEVLVPQGGFPFCDDPAPAGAGVNHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1080

DB 1020 EEVLVPQGGFPFCDDPAPAGAGVNHHRSSSTRSGGDLTLGLEPSEBEAPRSLAPSEG 1079  
 QY 1081 AGSDVFPDGLGMAAKGLQSLPTHDPSPLQRYSSDPVPLPSLETIDGVYAPLTCSPQREYV 1140  
 DB 1081 AGSDVFPDGLGMAAKGLQSLPTHDPSPLQRYSSDPVPLPSLETIDGVYAPLTCSPQREYV 1139  
 QY 1141 NQPDVPRQPSRPGPAPARPATATLER-----AKTLSPGKGVWQVDFAFGAVENPE 1195  
 DB 1141 NQPDVPRQPSRPGPAPARPATATLER-----AKTLSPGKGVWQVDFAFGAVENPE 1195  
 QY 1196 YLTPQGAAPQHPHPAPAFDNLVYWDQDPPERKAPPSCTFKCTPTAENPEYGLDVPV 1255  
 DB 1196 YLTPQGAAPQHPHPAPAFDNLVYWDQDPPERKAPPSCTFKCTPTAENPEYGLDVPV 1255  
 QY 1200 YLAPGRAPAPQHPHPAPAFDNLVYWDQDPPERKAPPSCTFKCTPTAENPEYGLDVPV 1259  
 DB 1200 YLAPGRAPAPQHPHPAPAFDNLVYWDQDPPERKAPPSCTFKCTPTAENPEYGLDVPV 1259

RESULT 2  
 Q90X70 PRELIMINARY; PRT; 1209 AA.  
 AC Q90X70;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor.  
 GN EGR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RX MEDLINE=9025888; PubMed=2342466;  
 RA Pech L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,  
 RA Earp H.S.;  
 RT "A truncated, secreted form of the epidermal growth factor receptor is  
 encoded by an alternatively spliced transcript in normal rat tissue."  
 RL Mol. Cell. Biol. 10:2973-2982(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RA Pech L.A.;  
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RA Guttridge K., Dawson T.L., Earp H.S.;  
 RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; M37294; AAF1408.1; -  
 DR HSSP; P11362; IFCK.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF01030; Recep\_L domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR Prodom; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; TyKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 KW ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 1209 AA; 134891 MW; 96FEETFCCLIB773 CRC64;

Query Match 45.5%; Score 3107; DB 11; Length 1209;  
 Best Local Similarity 49.8%; Pred. No. 1.8e-28;  
 Matches 636; Conservative 161; Mismatches 366; Indels 114; Gaps 26;

QY 3 LALCRMGLLALIPGA-ASTQVCTGTDMKRLPASPETHLDMRLHYGCCQVYQNL 61

```

Db      15 LAAALCAAG-----GALBEKKVCCGTSNRLTQCTFEDHFLSLORMFNCEVULGNLE 66
Qy      62 LITLPTNALSIFQDIQEVQGYLLIHNQROYPLRLRIIVRGTLQFEDNALVAVDNGD 121
Db      67 IIVYQENYDLSFKITIQEAVGYLLALNLTVERIPLENLQIIRGNALYENTVALAVLSN-- 124
Qy      122 PLNNTPVYASPGSLREQLRLTEILKGVYLQBNPOLCQODITLMDIHKHNNQAL 161
Db      125 -----YGTNKTGIRELPMRLQELLGAVRFSNNPLLCNWEITIQMRDIY-QDVFLSN 175
Qy      182 TLIDITNRS-RACHPCSPMKGSFRCWSESSEDCQSLTRTYCAGCA-RCKGRLPTDCHEQ 239
Db      176 MSMDVQRHLTGCPKCDPCSPNGSCMGSGEENCCKLTIKICAQCCSRGCRGSPDCCHNQ 235
Qy      240 CAAGCTGPKHSDCLACLHNHSGICELHCPALVTYNTDFFESNPNPEGTYTGASCYTA 299
Db      236 CAAGCTGPKHSDCLACLHNHSGICELHCPALVTYNTDFFESNPNPEGTYTGASCYTA 295
Qy      300 PNYVLTSDVSSCTLVCPLNQEVTAEDGTQCEKSCPCARCYGLMGQYIKANSFKTGI 359
Db      296 PRVYVLTSDVSSCTLVCPLNQEVTAEDGTQCEKSCPCARCYGLMGQYIKANSFKTGI 353
Qy      360 TELE-FAGCKKIFGSLAFPESEGDPPASNTAPLOEQQVEFTLEITGYLYISAMPDS 418
Db      354 TIKKPKYCTAISGDHLIPVAFKGSFTRTPPLPRELEITIKTYVEITIGFLLIQMBEN 413
Qy      419 LPLSYFONLOVYRGRILHNGAVSLTQGLGSMILRELESGSLALIHNNHLCFPH 478
Db      414 WTDLHAFELREITIRGTRKHGQPSLAVNLNTSLGRLKXISQDVYIISGNRLCYAN 473
Qy      479 TVPMDOLFENPHQALHTANREDECEVGEGLACHQICARHGMCGRPTQCVNCSQFLRQ 538
Db      474 TIKMKLFGTPNOKTKIMNRAKCKOKATNHCNPLCSSECGWGEPTDQVSCQVNSGR 533
Qy      539 ECVBERVYQGLPREVYVNRHCLPCHPECPONGSVTCGCPADQVACAHKDPFCYA 598
Db      534 ECVBERVYQGLPREVYVNRHCLPCHPECPONGSVTCGCPADQVACAHKDPFCYA 593
Qy      599 RCPGSKPDLSYMPIMKPFDEEGACQPCPINCTHSCVDLDDGCAEAGASPLTISFNN 657
Db      594 TPCSGIMGENNTL-VMKFADANNVCHLCHANCTYCAGAGLKGCC-QQPEGKITSITG 650
Qy      658 FIVSFVWIKYKVASHLERKROOKIRKTYMRLLQETLVEPLTSGAMPNQAQRILKE 717
Db      651 IVGGDLFTVVALGIGLFMRRLQVKKTLRLLOEREIVELTPSGAIPAQAHLRIKE 710
Qy      718 TELRKVYKLGSGAFGVYKGIWIPGENVYKIPVAIKVIRENSPKANEILIDEAYVAGV 777
Db      711 TELRKVYKLGSGAFGVYKGIWIPGENVYKIPVAIKVIRENSPKANEILIDEAYVAGV 770
Qy      778 GSPVYSRLIGICLTSTVOLVQLMPYGCILDHVRENRGLSGODLLNMCQIANKMSYLE 837
Db      771 DNPHVCRLLGICLTSTVOLVQLMPYGCILDHVRENRGLSGODLLNMCQIANKMSYLE 830
Qy      838 DVLVLRDLAARNVYKSPNHYKIDPGIARLDLDEHYADGQKVIKMMALESILRR 897
Db      831 DVLVLRDLAARNVYKSPNHYKIDPGIARLDLDEHYADGQKVIKMMALESILRR 890
Qy      898 RFTHOSDVMSYGVIVWELMTFGAKPYDGIPIAREIPDLLEKSGRLPQPICTIDYVMWV 957
Db      891 IYTHOSDVMSYGVIVWELMTFGAKPYDGIPIAREIPDLLEKSGRLPQPICTIDYVMWV 950
Qy      958 CWMIDSEGRPRELVESEFSKARPPORFVVIQ-NEDIGPASPLDSTYRSLLEDDMGD 1016
Db      951 CWMIDSEGRPRELVESEFSKARPPORFVVIQ-NEDIGPASPLDSTYRSLLEDDMGD 1010
Qy      1017 LVDAEYLVPOOGFPCPDPAFAGAGVHHRSSSTRGCGDGLTGLSESEENPRSLA 1076
Db      1011 VVDADEVYLIPQGGF-----NSBT-----SRPBL 1036
Qy      1077 PSEBAGSDVFPDGLCMGAKLQSLFTHDPSPLQRYSDPTVPPSEH--DGYVAPLTCS 1134
Db      1037 SSLSANSN-----SSTVACINENGSCRVKEDAFLORYSSDPTSVLTEDNIDDTFL----- 1086

```

```

Qy      1135 PPEYVNPDPVPPPPRECPPLPAPAGATLERAKTLISGCKGKGVYXVFAFGAVNP 1194
Db      1087 PVPEYVNP-SVXPKPAGSVQNPVYHNNGLHF-----AGRDILHYQN--PHSAVSNP 1135
Qy      1195 EYL-TPQGAAPQHPHPPAPFAPFDNLVYMQ-----DP-----PERGARPSFTF 1237
Db      1136 EYLTAQ-----PTCLSSGFPDSSALWIKGSHQMSLDNPYQODFFPEAKPNCIF 1186
Qy      1238 KGTFAENPEYGLDVP 1254
Db      1187 KG-PTAENAEYLRVAP 1202

RESULT 3
Q9EP98
AC Q9EP98: PRELIMINARY; PRT, 1210 AA.
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Epidermal growth factor receptor isoform 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearall R.S., Green P.J., Yee D., Lampand A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Scheel C., Pearall R.S., Green P.J., Yee D., Lampand A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RP [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Scheel C., Pearall R.S., Green P.J., Yee D., Lampand A.L.,
RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maible N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egfr transcripts encoding truncated receptor
RT isoforms";
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.

```

KW ATP-binding; Receptor; Transferase.  
 SQ SEQUENCE 1210 AA; 134840 MM; 62CD021C9DE32E18 CRC64;  
 Query Match 45.0%; Score 3078; DB 11; Length 1210;  
 Best Local Similarity 49.3%; Pred. No. 3e-226;  
 Matches 627; Conservative 163; Mismatches 372; Indels 110; Gaps 23;

11 LLLALLPGLA--STVCTGTDMLKLPASPTHTDMLRLHLYGCGVQVGNDELTYLPTN 68  
 14 LLLALCAAGAALEKKVCCOTSNRLTQLGTPEHFLSLGRMNNCEVVLGNLEITYQGN 73  
 69 ASLSFLODIOEVQGYVLIHANOVYPLQRLKIVGTQVPEENYALAVLDNDGPLNNTTP 128  
 74 YDLSFLKTIQEVAGVYLIANTVERIPLENQIRGNALYENTYALALISN----- 124  
 129 VTGASPGIRELOLRSLTEILKGVLIQRPQVCYDITLMKDI----FHKNQDLATLI 184  
 125 -YGNRTGLRELPMRLQELIGAVFNNPILCNMDITQMRDIYQVNMNSMDL---- 180  
 185 DTRSRACHPCSPCKSGRCWGBESSDQSLRTVCAGCA-CKGPLEPTDCCHQCAAG 243  
 181 -OSHPSCKPCDPSCKNGSCWGGGECNCKLTKILAQOCSHRCGRSPSDCHNQCAAG 239  
 244 CTGPKHSDCLACHENHSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTAQPNY 303  
 240 CTGPRSDCLVCKFQDEATCKDTCPPLMLNPTTYQMDVNEBKISFGATCVKCKPNY 299  
 304 LSTDVSGCTLVCELNQEVYTAEDGTGRCCKSPKPCAVCYGLMOYIKANSKFIGITBLE 363  
 300 VVTDHGSVYACGPDYEV-EDGIRCKCKCDGPKCKVNGIGIGIFK-DLTSINATNIK 357  
 364 -FAGCKKIFGSLAFIESPDPDASNTAFLOEQVFTLEIGLYLISMPSLDPL 422  
 358 HEFYCTAISGDHLILVAFKGSFTTPPLPRELEITLVEITGFLIQMPNNMTDL 417  
 423 SVFQNTQVIRGRILHNGAVSLTQGLISMLGRSLREXSGSLALHNHTLCFHYTPW 482  
 418 HAFENLEIRGRKQGGFSLAVGNTSLGRSLKESDQVLIISGRNLCYANTINW 477  
 483 DQLEFNPHQALHTANRPEDECVGSLACHQUCARHCHGPGPTQVNCQPLRQECVE 542  
 478 KKLFGTPNOKTKIMNNRAKDCAVNVNCPICSSGCGPREDVSCONVSRRECE 537  
 543 ECEVLQGLPREVYNAHCLPCHPECPONGSVTCFSPADQCAVHYDPFCVACPS 602  
 538 KCVILGEPREFEENECIQCHPECPQAMNITCTGRDNCIQCHAYIDGHCVKTCDA 597  
 603 GVKPDLSTYMPKFPDEGACQCPPICTHSCVDLDDKCCPABQRAPLTSIFNNFTVSF 662  
 598 GINGENNTL-VMKYADANNVCHLCHANCTYGCAGPGLQCEVWPSGPKIPSIATGVGL 656  
 663 WLRVPKVASHLEKROQKIRKVTMRLLQETELVEPLPSGAMPQQAOMRLIKETELK 722  
 657 LFLVVALGIGLFMRNRHIVRKRLRLQERLEVEPLPSEAPQAQALRLIKETEKK 716  
 723 VKVLGSGAGTYVKGIMIPDGENVKIPVAIKVLENTSPKANKELIDEAYVAGYSPV 782  
 717 IKVLGSGAGTYVKGIMIPDEGEVKIPVAIKELRENTSPKANKELIDEAYVAGYSPV 776  
 783 SRLGICLSTVQLVQLMPYGLDHNENGRGLSQDLLWCMQIANGSYLEDVRLY 842  
 777 CRLGICLSTVQLVQLMPYGLDHNENGRGLSQDLLWCMQIANGSYLEDVRLY 836  
 843 HEDLAARNLVKSPNHVKTIDFGLALDIDETEHADGKVPKIMMALESLIRRFHQ 902  
 837 HEDLAARNLVKTPQHVKTIDFGLALDIDETEHADGKVPKIMMALESLIRRFHQ 896  
 903 SDVWSYGVTVWELMTGAKPYDGIAPRRLPDLLEKGERLPPICTIDVYMIYKCMID 962  
 897 SDVWSYGVTVWELMTGSKPYDGIAPRRLPDLLEKGERLPPICTIDVYMIYKCMID 956  
 963 SECRPFREIYSEFSMARDPQRFVYIQ-NEDLGPASPLDSFFYSLEDDDMGLVAAE 1021

DB 957 ADSRPFREIYSEFSMARDPQRFVYIQ-NEDLGPASPLDSFFYSLEDDDMGLVAAE 10216  
 QY 1022 EYLVPGCGFPCPPAPAGAGVHHRRSSSTRSGGDLTLGLPSEEEAPRSLAPSEGA 1081  
 DB 1017 EYLVPGCGF-----NSPST-----SRTPLSSLSA 1042  
 QY 1082 GSDVFDLGLGAAKGLQSLPTHPSPLOQYSEDPVLPSET--DGYPALTCSPQPEY 1139  
 DB 1043 TSN-----NSTVACINRNGSCVKEADALQRYSSDPGTAVMEDNIDDAFL-----PVPEY 1092  
 QY 1140 VNQPDYVQPPSPREGLPAPAPAGATLEAKTILSPKNGVYKDVFAFGAVENPEYL-T 1198  
 DB 1093 VNQ-SVPRPAGSVQNVYHNQPLP-----APGRDLHYQN--PHSNVAGNPEYLNT 1141  
 QY 1199 POGGAAPQHPPPAFSPAFNLVYWDQ-----DP-----PERGAPPTSTFKGPT 1242  
 DB 1142 AQ-----PTLSSGGRFNSPALMIKSGHNSLNDPVDQDQFFPRKTRNGIFKQ-PT 1191  
 QY 1243 AENPEYGLDVP 1254  
 DB 1192 AENAEYLRVAP 1203

RESULT 4  
 Q9YH40 PRELIMINARY; PRT; 1165 AA.  
 AC Q9YH40;  
 DT 01-MAY-1999 (TREMblrel, 10, Created)  
 DT 01-OCT-2000 (TREMblrel, 15, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel, 21, Last annotation update)  
 DE Receptor tyrosine kinase proto-oncogene.  
 GN XMRK.  
 OS Xiphophorus xiphidium.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 CC Acanthomorpha; Acanthopterygii; Percomorphi; Atherinomorpha;  
 CC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
 NX NCBI\_TaxID=8086;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RX MEDLINE=96241172; PubMed=9582016;  
 RA Dinterjefevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,  
 RA Altschmid J., Scharl M.,  
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by  
 RT overexpression and mutational alterations.";  
 RL Oncogene 16:1681-1690(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RA Scharl M.,  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; U53471; AAD10500.2; -.  
 DR HSSP; P11362; IFCG.  
 DR InterPro; IPR000345; CyC heme bind.  
 DR InterPro; IPR000494; EGFR L domain.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR002174; Eurin-like.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR001245; Tyr\_Pkinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF01030; Recep\_L domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR PRODOM; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; TyKc; 1.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN 1.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS0011; PROTEIN\_KINASE\_DCM; 2.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.

SEQUENCE 1165 AA: 129614 MW: 777E38D8771A74E CRC64;

Query Match 39.3%; Score 2685; DB 13; Length 1165;  
 Best Local Similarity 45.3%; Pred. No. 3.2e-196;  
 Matches 579; Conservative 161; Mismatches 386; Indels 152; Gaps 32;

1 MELALCRWGLLALPPG-AAST-----OVCTGTMKRLPASPETHLMRLHLYQGQCV 55  
 4 LELIEL-----LHLLLSIRCCSTDPDRKVCQGTSNQMTM---LDNHYLKKMKMSGCNV 56  
 56 VQGNLELLYLPNTNASLPLQDIOEVGVYVLAHNOVQVPLQRLIRVSTQLFEDNYALA 115  
 57 VLENLEITYEOENODLSFLOISIOEVGVYVLAHNEVSTIPLVNLRLIRQNLYEENFTLL 116  
 116 VLDNGPLNNTPTVYGASPGGLRELQSLTELKSGVLIQGNPOLCYDTLLKMDIRHK 175  
 117 VMSNYOK-PPSSP--DVYVGLKQLOLSNLTLEILSGGVKSHNPLCNVETIIMWDIDYK 173  
 176 NNQLALTLIDTRSRACHPCSPCKSGRCMGSESDCQSLTRTVAGGC-ARCKGRLPTD 234  
 174 TSNPTMNLIPHAFFERCCQKCDPGCVNGSCWARGPHGCKFTLLCAQCNRCRCRKPPID 233  
 235 CCHQCCAGCTGPKASDCLACHFNHSGICELHCPALVYNTDTESMPNPGRYTFGAS 294  
 234 CCHQCCAGCTGPRATDCLACRD FNDGTCCKDTCPPKIYDIVSHQVNDNPKIKYTFGAA 293  
 295 CVTACPRNYLSTDVSGCTLVCPLNQEVTAEDGTQRCESKPCARCYGLM-----QYI 350  
 294 CKEGPSNYVTE-GACVASCAGMLEVD-ENGKASCRCPCDVCRCVCDGIGISLMTI 351  
 351 KANSKFIGITTELFAGCKKIFGSLAFLESPFGDPSANTAPLOPELOQVFETLEETGYL 410  
 352 AVNSTNIG---SFSNCTKINGDILNRSFEGDPHYKIGPMDEPHLMLTIVKTIITYL 407  
 411 YISAMPDLPDLSVQNLQVITGRILHNGAYS-LTLOGIGISWLGIRISREIGSLALIH 469  
 408 VLMMPENNTSLSVQNLLEIRGRITFSRGSFVAVVQVSHLOMTLSLSKEVASANVILK 467  
 470 HNTLCFVHTVPMDOLEFRNPQALHTARPEDECVGSLACHOLCARGHCGPPTQCV 529  
 468 NTPOLRYASTIMWRLLFSEDSIETDART-----ENQTCNNCSDEDCMGPPPTCV 520  
 530 NCSQFLRGQECVECRVLQGLFREYVNAHCLPHEPCOPONGSVTCFGEPAQCVACAH 589  
 521 SLHVRGRCVASCNLLQGEFREQVQDGRQVQCHQECILVQDLSITCYGPPRANSCKAH 580  
 590 YDPPFCVACPCSPGVKPPDLSYMPWKPFDEBAGACQPCINCHSGVDDDDGCAAE--QR 647  
 581 FODGPOCTPRCHGMLSDGDTL-IMKYADKMGQCCPCHONCTQGSGLSGCRGDIVSH 639  
 648 ASPLTSIFNNFTVSFMLEVPKVASASLLEKRRQOKIRKXTMRLLQETELVEPLPNSGAP 707  
 640 SSLAVGLVSGLLITIVALLIV--LRRRRIRK-RKRTIRRLQEKELVEPLPNSGAP 695  
 708 NOAQWRILKETELERKYLKVGASGFTVYKGMIPGSENVKIPVAIKULRENTSPKANEI 767  
 696 NOAFRIILKETELERKYLKVGASGFTVYKGMIPGSENVKIPVAIKULRENTSPKANEI 755  
 768 LDEAYVAGVSPYVSRLLGICLSTVQVLTQMPYGLDLHVENRGRLSODLLNMC 827  
 756 LDEAYVMSVDHPCRLGLGICLSTVQVLTQMPYGLDLHVENRGRLSODLLNMC 815  
 828 CIAKMSYLEDVRLVHRDLAARNLVKSPNHYKTTDFGLARLLDIDETEHYADGKVPYIK 887  
 816 CIAKMSYLEDVRLVHRDLAARNLVKSPNHYKTTDFGLARLLDIDETEHYADGKVPYIK 875  
 888 WMALESILRRFTSHQSDVMSYGVTVWELMTGAKPYDGIAPREIIPDLLEKERLPPOPIC 947  
 876 WMALESILQWYTHQSDVMSYGVTVWELMTGSKPYDIPAKELASVLENERLIPPOPIC 935  
 948 TIDVTMIVKCMIDSECRPRRELIVSEFSRMAPDQRFVVIQNEIDGLPASPLDSTYRS 1007  
 936 TIEVMTILKCMIDPSSRPRRELIVSEFSQWARDPSRYLVIGQ---NLPSPSDRILFSR 992

1008 LLEDDMGDLVDAEEYVPPQGFPCPDPAAGAGVHRRHSSSTRGGDLTLGLPSE 1067  
 993 LLSSDD--DVVDADEYLL-----RYKRLN-ROGS----- 1018  
 1068 EEARSPPLASEGAGSDVFGDDLGMAKALQSLPTDPSPLQYSEDPTV-PLPSETDG 1126  
 1019 -----EPCIFPNHG-----PVRENSTALRYISDPTQNALKEKLDG 1053  
 1127 YVAPLTCSPQPEYVNOVDVPPQ-----PSPRE-----GPLP-AARPAATLERAKTL 1173  
 1054 H-----EYVNGGSESSRLSDIYNPNVEDLIDMGCPVSLSSQEAFTNSREYTL 1103  
 1174 SPGNQGVKQVFAFGAVENPEYITPPGGAAPQCHPPAPSPAFDNLVYWDQDPERGAP 1233  
 1104 NTNQNSL---FLVSSGMDPDY---QAG-----YQAAF-----LPQTGAL 1138  
 1234 PSTFKGPTAENPEYLG 1251  
 1139 TGNMGFLPAENLEYLGL 1156

RESULT 5  
 ID Q9W6F6 PRELIMINARY; PRT: 1137 AA.  
 AC Q9W6F6  
 DT 01-NOV-1999 (TREMBLER, 12, Created)  
 DT 01-NOV-1999 (TREMBLER, 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBLER, 21, Last annotation update)  
 DE Receptor tyrosine kinase (Fragment).  
 GN ERBB4.  
 OS Gallus gallus (chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HINDERAIN;  
 RX MEDLINE=99263203; PubMed=10328884;  
 RA Dixon M., Lumsden A.;  
 RT "Distribution of neurotrophin-1 (ng1) and erbB4 transcripts in  
 embryonic chick hindbrain.";  
 RL Mol. Cell. Neurosci. 13:237-258(1999).  
 DR EMBL, AF121963; AAD31764.1; .  
 DR HSBP, P11362; IFCX.  
 DR InterPro: IPR000494; EGFR\_L domain.  
 DR InterPro: IPR000719; Euk\_kinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001368; TNFR\_c6.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF01030; Recep\_L domain; 1.  
 DR Pfam: PF02757; YLP; 2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_kinase; 1.  
 DR SMART: SMO0261; FY; 3.  
 DR SMART: SMO0219; TYKc; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.  
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.  
 DR PROSITE: PS00652; TNFR\_NGFR\_1; UNKNOWN\_1.  
 KM Kinase; Tyrosine-protein kinase.  
 FT NON\_TPR 1  
 SQ SEQUENCE 1137 AA: 127927 MW: 4D616436F87DC84F CRC64;

Query Match 38.7%; Score 2646.5; DB 13; Length 1137;  
 Best Local Similarity 46.5%; Pred. No. 2.7e-193;  
 Matches 532; Conservative 167; Mismatches 357; Indels 89; Gaps 26;

161 LCTQDITLTKMDIFHKNNQALTLIDTRSRACHPCSPCKSGRCMGSESDCQSLTRTV 220

```

Db 3 LCFADITHMDIYRBNWASNTLVPTNGSSGCGRCKSKCTG-RCMPTENHQCITLTKYTC 61
QY 221 AGGC-ARCGPLPTDCCHEGCAAGCTGPRHSDCLAFHNSGICELHCPALVTYNTDTF 279
Db 62 AECQDRCYGPVYSDCHRECAAGCGSPKDTDFACMNFNDSCACTOCPCPTFFVNPPTTF 121
QY 280 ESMNPDEGRYTFGASCVTACPRVYLTSDVSCCTLVCPRLNHOEYTAEDTQRCCKSKPCA 339
Db 122 QLEHNNAKTYTGAFCVKKCPHNFV-VDSSSCVACAPSSKMEV-BENGIMCKPCTDTCF 179
QY 340 RVCYGLGMOYIKANSKFIGITTELE-FAGCKIFGSUAFIPESFDGDPASNTAPLOEBOLO 398
Db 180 KACDGIQTGSL-USAQTVDSNNDKFNCTKINGNIIPIVTHGHPHYTHIAINPEKLN 238
QY 339 VFETLEITGYLISAMPBLSPLSVFQNLQVIRGLILNGANSLTLOGLISWIGRL 458
Db 239 IFQVREITGYLISQWPEMNTDFRVFSNLVTIGSALYSGLSLILKQOGITSLQFQL 298
QY 459 RELSGGLALIHNTLHCFVHTVWDOLFERNHQAALHTANRPEDECVGEGLACHOLCARG 518
Db 299 KQISAGNIYITDMSNCYHTVNTSLFSTPSQKVIHNNKKAENCTADGWCNLCSSD 358
QY 519 HCGPPTQCVNCSQPLRGECEVEGRVLOGLPREVNAHCLPCHPEQOP-QNGSVTF 577
Db 359 GCGPDPDCLSKREIRGRTICIESCNLYDGEFREANSGVCECDPQCEKMDNWTGY 418
QY 578 GPADOCVACAHKDPFPCVAPRSPGVKPDLSYPIKPFDEGACQPCPINTGSCVDL 637
Db 419 GPGPDHCTCFHFKDGNCEKCPDLOQANSF-IFKADEBRECHPHNCTQOC--- 473
QY 638 DDGXCPEBORA--SPLTSIFNNFTVSFWLRVPKVAS-----HLEKRRQ 680
Db 474 --RG-FASHDCIYPTW---RSTLPOHATPPLIAGVIGLFIIVIMGLTFAVYVRKS 527
QY 681 KIRKYMRLLOTELEVEPLTPSGAMPNOQMRLIETELRKVYLGSAFGVYKGI 740
Db 528 IKKRALRRFL-ETELVEPLTPSGTAPNOQMLKILETELKRVKLGSAFGVYKGIW 586
QY 741 PDGENVKIPVAIKYLENTSPKANKELDEAVYVAGVSGPYVRLIGICLSTVQLVTL 800
Db 587 PEGEIVKIPVAIKILENTSPKANKELDEAVYVAGVSGPYVRLIGICLSTVQLVTL 646
QY 801 MRYGCLLDHYENRGLSGODILNMCQIAKMSYLEDVRLVHDLAANVLKSNHIX 860
Db 647 MRYGCLLDHYENRGLSGODILNMCQIAKMSYLEDVRLVHDLAANVLKSNHIX 706
QY 861 ITDFGLARLDIDETEYHADGKVPKIMMALESILRRFTHQSDVMSYGVTWELMTFGA 920
Db 707 ITDFGLARLDIDETEYHADGKVPKIMMALESILRRFTHQSDVMSYGVTWELMTFGA 766
QY 921 KPYDGIPTAREIPULLEKGERLPORPCTIDVYIMYKCMKIDSECPREELVSESRMA 980
Db 767 KPYDGIPTAREIPULLEKGERLPORPCTIDVYIMYKCMKIDSECPREELVSESRMA 826
QY 981 RDQORFVIONED-LBPASPLDSTFYRSLLEDMDMDLVDAEYVLPQOGFCPPDAPGA 1039
Db 827 RDQORFVIONED-LBPASPLDSTFYRSLLEDMDMDLVDAEYVLPQOGFCPPDAPGA 885
QY 1040 GGVVHRRHSSSTRSGGGLITGLLEPSEEBAPRS--PLAP-SEGAGSDVFDGDLGMAAK 1096
Db 886 RTRIDSNRNQFVYRDGGAFAEGV-PMYRAFPGLIPEARVAGAAFAEFEDTCCGTLR 944
QY 1097 GLGSLPTDPSPIQRSEDPVPLPS-----ETDGYVAPLTCSPQRPVYVQDPVRP 1149
Db 945 KQVATLAKEDSSTQRIASDPTVFIPIRVIRGELDEDEGYVTPMDKDKTYLNPVEENPV 1004
QY 1150 PSREGEPLPA-RPAGATLERAKTISPKNVYKVF-----AFGAVENPEYLTPO 1200
Db 1005 SRKNGDLQAVNDPEYHN-----APNGQPKADEYVNEPLVNTFAUTLEAEYL--- 1054
QY 1201 GGAAPQHPPPAPSPAFDNIYYWDQDPERGA--PSTFKGTPT-----AE 1244

```

```

Db 1055 -----KNULPEKAKKAFQNDPDMYKNSLPRETTQHPDVLQESTYFYKQNGRIPIVAE 1109
QY 1245 NPEYL 1249
Db 1110 NPEYL 1114

RESULT 6
P79754 PRELIMINARY; PRT; 1328 AA.
AC P79754;
DT 01-MAY-1997 (TRENBLrel. 03, Created)
DT 01-MAY-1997 (TRENBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)
DE Erib3.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostei;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
NCBI_TaxID=31033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99177347; PubMed=10077531;
RA Gelliner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
rubripes."
RL Genome Res. 9:251-258(1999).
RT EMBL; AF056116; AAC34391.1;
DR HSSP; P11362; IFCX.
DR InterPro; IPR00494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF000257; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Prodom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS0011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase
SQ SEQUENCE 1328 AA; 148613 MW; A3330393258B647E9 CRC64;

Query Match 33.1%; Score 2260; DB 13; Length 1328;
Best Local Similarity 39.9%; Pred. No. 1.2e-163;
Matches 516; Conservative 155; Mismatches 418; Indels 204; Gaps 33;

9 WGLIALLP--GAASDQ---VCTGDMKRLPASPEPTHLDMLRHLYQGCVYQNL 62
Db 4 WRLIMCVASRLRAASQTOEAVCPGTONGLSSTGSENOYLNKDRYKGEIIMGNLEI 63
QY 63 TYLPTNASTLSTLODIOEYQGYVLIHNVORVPLORIRIVYGTOLFEDNYALAVLNDGP 122
Db 64 TOIESNMFSLTKIRLETGYVLIAMNHFOGILPQGLRVITGNSLYERRFLSFLN--- 120
QY 123 LNTTPTVTGASPGGLRELQRLSLTEILKGVLIQRPOLCYQDTILMKDIFKXNQLALT 182
Db 121 ----YFKDG--PSGLNQLGLMNLTEILLDGGVQIINXKRLRGVPMVYRDDI--RNNDAPIE 173
QY 183 LIDTRBSAACHPCSPMCGSRCHGESSSDQSLTRTYCAGGC-ARCGPLPTDCCHECA 241
Db 174 IQNGERGVCH--KSC-GNTCMWPGKDCQILTKIVCAPOQNDRCFTGSPRDCHECA 229
QY 242 AGCTGPKHSDCLAFHNSGICELHCPALVTYNTDTFESMNPDEGRYTFGASCVTACPY 301
Db 230 AGCGPLDPTDCAFRLFNDSGACVPCQGLIYKQFQWETNNAKYQSGISCVSCPT 289
QY 302 NYLSTDVSCCTLVCPRLNHOEYTAEDTQRCCKSKPCARCYGLGMOYIKANSKFIGIT 360
Db 290 HFV-VDGSSCVSVCPDRMEV--ERGSQRCQELCSGLCPKVCETGTGAE---QRTVDSS 342

```

```

QY 361 ELE-FAGCKKIFGSLAFLPESFDGASNTAPLQPIQLQVFFETLEITGLVYISAMPDSL 419
DB 343 NIDSFNCTKIQSLHFLVTLGIDPCKVPPDADKLEFVRVITITILNTQSPKEL 402
QY 420 PDLSEFQNLQVIRGRIILHNGAVSLTLQGLISWLGRLSRELGSGALILHNTHLQCFVHT 479
DB 403 NDLVSFSSLTIGRSLFKRFLSMWRIRPILTLISGLRSLREISDGSVYISQNAHLCHHT 462
QY 480 VPMDOLEFRNH-QALLHTANRPEDECVGSLACHQICAGHCHMGPEPTCCVNCQFLRQ 538
DB 463 VMTQLFRGSRVANSILNSRPMACVADGRVCDPLCSGSGCGPDDCLSCRNYSRHG 522
QY 539 ECVEECRVLQGLFREYVNAARH-CLPCHPEQOPONGSVTCFEPADOCVACAHYKDPFCV 597
DB 523 TCYAGCHFNISGIRREFRAGLNGVACHPECKPOTGASCTGPEADCMCTFRDGPYCM 582
QY 598 ARCPGVKPDLSMPTWKPDEEGACQPCPINCTHSCVLDLDDKCPAEGRAS----- 649
DB 583 SSCPAGVN-DGEKGLIFKFPNREGHCPCQONCTQCGSGPLNDCLARLTSSQITG 641
QY 650 -----PLTSIFNNFTVSFMLRVKVASHLKRRQOKIRKYMRLLOTELVEPLTBSG 704
DB 642 IALGVAGLIF-CLVLFPLGM-----LYHGLAIRRRKRRRRLSSESFEPLGP-G 691
QY 705 AMPNOQMRILKTELKRVKVLGSGAFGVYKGIWIPDENYKIPVALKRENTSPKXAN 764
DB 692 EKGTKVHAHLKRSDLRKIKPLGSGVGTGSKGFWPEDETKIPALITIDSSGRQIF 751
QY 765 KELDEAYVMAGVSGSVYRLLIGICTSTVOLVTOCMYGCILLDHYRENRGLSGDILN 824
DB 752 TETTHLLMSGSDHPIYIRLLIGICPTGLQVLTQSSGSLLEHTRQKHTLDJQRLIN 811
QY 825 MCMQIAKMSYLEVDVLRDLAARVNLVKSNNHVKITDFGARLIDIDETRYHDDGKV 884
DB 812 KCVQIAKMYIIEHNHVAKNLAARHILKANDYQOISYGAIDLIPDKKYVSETKT 871
QY 885 PIKMALESILRRFTHQSDVMSYGVTTWELMTFGAKPYDGIAPAEIPDLLEKGRLP 944
DB 872 PIKMALESILRRFTHQSDVMSYGVTTWELMTFGAKPYDGIAPAEIPDLLEKGRLP 931
QY 945 PCTIDVYMWKCMWIDSECPREIIVSEFSRMAPRQRRVVIQNEGLGASLDSTF 1004
DB 932 ALCTIDVYMWKCMWIDSECPREIIVSEFSRMAPRQRRVVIQNEGLGASLDSTF 980
QY 1005 YRSLLEDDMDGLVAEEYLPQGGFPCDPAFGAGVHNRHSSSTRSGGDLTGLE 1064
DB 981 -----EDSGMGEL-----RGSER-----GLEADLE 1003
QY 1065 PSEERARSRPLAPSEGASDVFDGLGMC---AAGGLQSLPTHDSPLQ-----R 1111
DB 1004 EDEEB-----GLGDRFATPSLOPSPSMSTSPQINSYVMTOLR 1042
QY 1112 YSEDPVPLPSETDGVAPLTCSPQPOP-EYVNO-----PVRRPQPSRRGP 1156
DB 1043 VD-----FVSGQGHIGVLPMSPSVDITIRQLMWQRSLSSVRLTPDSAFRRSRABE 1096
QY 1157 L--PAARPAATLEAKTILSPGKNGVVDVAFGAIVENPEYLTQGGAAAPHPAPAS 1214
DB 1097 LCEDGAQCGAIFRRV-----FGSERGV-----PQGG----- 1122
QY 1215 PAFDNLVYWDODPERGAPPTFGKPTAENPE 1247
DB 1123 -----QQRKLTASSPSSFKTWADEDE 1146

```

## RESULT 7

```

ID Q9BIH9 PRELIMINARY: PRT: 1433 AA.
AC Q9BIH9;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).

```

```

GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_Taxid=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SLA.
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor ";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ01655; CAC35008.1; ..
DR HSP; P11362; IFGK.
DR InterPro; IPR000345; Cytc_heme_bind.
DR InterPro; IPR00494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_Pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_thr_Pkinase.
DR InterPro; IPR001245; Tyr_Pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Pkinase; 1.
DR PRINTS; PR00109; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00261; FU; 7.
DR SMART; SM00220; S_TKC; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00101; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KM Receptor.
FT NON_TER
SQ SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

```

```

Query Match 29.1%; Score 1988.5; DB 5; Length 1433;
Best Local Similarity 32.4%; Pred. No. 8.2e-143;
Matches 469; Conservative 196; Mismatches 397; Indels 385; Gaps 38;

```

```

QY 26 CTGDMKRLRPASPEHLDMLRHLVYQCGVQVNLVPTNLSFLDIDQVQVYL 85
DB 1 CIGTNGMSVYANNEYHYKULRDRYNTCTVDSGLTWTQNIIDNLFQHIRVYTVL 60
QY 86 IAHNOVQVPLQRLIRVGTOLF-----EDNYALAVLDNGDPLNNTTPVTCASGGLREL 140
DB 61 ISLYDLRQVILLPRQIIRGRITFTFKLNKEAYGLFV-----SFSHMTL 104
QY 141 QLRSLTEILKGVILQGNPOLCYODTILMKDI-FHKNNOLATLIDTNRSPAGCPSPMC 199
DB 105 ELRALRDIILGSGVGFNNYNLCHMKSLNMBEILAPOTSMQYTFNESPFRVCPCHPSC 164
QY 200 KGSFCWSESSDCSLTRTVACAGCA--RCKGPLPTDCHEQCAAGCTGKPSDCLALH 257
DB 165 EVG-CWEGEGAHNCQRFKLNCSQCSGRCFGFPRBCHLFCAGGCTGTQSDCLACKN 223
QY 258 FNSHGICELCPALVTNTDTPESMPYBERRYFGASCYACAPNYLSTVGSCTLVCLP 317
DB 224 FYDDGVCKQCPQWQIATNPNTNFWBPBDDKIVAGATCAKCP-EHLKNGACVRRCKP 282
QY 318 HNOEVTAEADGTORCEKSCPCARVVCYAGIYKANSKFIGITELFEAGCKTIFGSLAF 377
DB 283 GKMPQNSE-----CVPEKGVCPKCPCEGI-----VHSDNIG-----NYKDCITIEGSLIEL 329
QY 378 PSEFQDGPASNT-----APLQPELOVFFELTBETGLYVYISAMPDSLPDLSVQNLQ 429
DB 330 DQSFQDQVYTNFSPGPRYIKIDPRLVFTSVKELTGFINIQAHHPNFTLTNYFRNIE 389
QY 430 VIRGRIHNGAV-SLTLOGIGISWLGRLSRELGSGALILHNTHLQCFVATVPMDOLEFR 488
DB 390 VVGGRQIKENLFAVSVYIVKTSLSLEKSLKRVNSGSIYLENSDLCFVEDIDMSEIKKS 449

```

```

QY 489 PHQALHTANREDECEVGEGLACHOLACGHCWGPPTQCVNCSQFLRGQCEVEGRVLQ 548
DB 450 SDHEVWVQKRNATTECHHEEMECSEGCSSAGCKGKPEQCLEKRVKXKGLDCK-- 506
QY 549 GLPREY-VNARHCLPHGPCQFONGSVTCGPEADCCVACAHYKPPFCVACAP----- 601
DB 507 SLERLYSVDSKTDGCHQOECKD-----FCYGNEDCGSCMVKDRCFVACCPPTTKAM 561
QY 602 -----SGVPEDSLVSVMKFPD----- 618
DB 562 NGCTCNCHKTVCGRGPRDTIAPDGCISCDAKIGSDAKIERCLMDESCPDGYSIDVL 621
QY 619 -EEG----- 621
DB 622 QEEGPLKQSGKAVCRCKCRCKCTGYGFHEQFCQECTGYKKGEQCEDECFQDYANBE 681
QY 622 --ACQPCINCT-----HSCVDL-----DD-----KCCPBE-----QPA 648
DB 682 TRICLPCHQECRGCHGLGDDHHCENLKLFEQDPYDNTFTTCVSNCPKSHPKRPPQEA 741
QY 649 SPLTSIFNFTVSFWLRY-PKVS-----ASHLEKRRQOKIRKXTM 687
DB 742 GKICPYCSADSWGSGRIEFGQVKTVMGSMALLLCVFGIAFLFGRHKKKDAVAK 801
QY 688 RLLQETELVEPTSGAMPNQAMRLKTELKXKXVJGSGAFGYVYKIMTPDGENK 747
DB 802 TMLACCESEPRPSPVNGVNLTKRIKEAETIRGCVGMGFGVFGVWMPGEGSVK 861
QY 748 IPVAIKVIRENTSPKANKETIDEAAYMAGVSPYRLLIGICTSTVQVLTQMPGCLL 807
DB 862 IPVAIKVIRENTSPKANKETIDEAAYMAGVSPYRLLIGICTSTVQVLTQMPGCLL 921
QY 808 DHVRENKRGISGODLLNMCQIAKGSYLEDVFLVHDLAARNVLKSPNHYKITDGLA 867
DB 922 DYVRNKKKXIGSALLNWSQIARGMAYLEERLVRDLAARNVLQTPSCVXITVGLA 981
QY 868 RLIDIDETEYHAGGVPIKMALESILRRFTHSDVSYGVTVMLTQPKKPDGJP 927
DB 982 KLDLFDSDERAGGKMPKIMLBECHRVFTSKSDVWFGITIMELLTYGARPYENP 1041
QY 928 AREPIDLEKGERLPOPPICITIDVIMVCMWIDECRPRELVSERSMARDPQREV 987
DB 1042 AKDVPELIEIGHKLPQDIDCSLDVYCLISCVLADARPTFKQLAETFAEKARDGRL 1101
QY 988 VIONEDUGRASPLDSTYRSLLEDDMDGLV----- 1018
DB 1102 MI-----PDGKFWRLPSYTNODEKDLITLAVMAAAAAAGASNVVDSITIA 1152
QY 1019 DAEYIVPQCGFCPPDPAPAGAGVHHRSSSTRSGGDLTLGLPSEEEAPRS----- 1073
DB 1153 EIDEXILOPKTRPSIMLPSPA-----VERPS-DEMKSLIRYCK 1188
QY 1074 -PLAP---SEGASDVFDGDLGMAKGLQSLPTHDPSPLOKRYSEDPVPLPSETDGYVA 1129
DB 1189 DPLKPDDETGHGKEV-----GVGGIR-----INTLPLDEDDYLM 1222
QY 1130 PLTCSPOPEYVNOVDVPPPPSPREGFLPAPAPAGATLERAKTILSPGKXGVVDVAFGG 1189
DB 1223 P-TCQO---NOS-----TBO---YMDLIGVTA 1243
QY 1190 AVENPEYL-----TPOGGAPOPHPPAPSPAFDNLVWDQDPERGAPSTFKCT 1240
DB 1244 SVNPPEYLMGSTOAIAGLAGSWG--PHTP-----PNTNPM 1280
QY 1241 PTAENPE 1247
DB 1281 PTHQHSQ 1287

```

RESULT 8  
 ID Q9UK79 PRELIMINARY; PRT; 419 AA.  
 AC Q9UK79;

```

DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RL autoinhibitor.";
RN [2]
RP Proc. Natl. Acad. Sci. U.S.A. 96:10669-10874(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Hennen W.D.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF177761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; FCIER347E2D030C CRC64;

Query Match 27.4%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 1.2e-134;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELALCRWGLLALLPPGAASVQCTGTDMKRLPASPEHLDMLRLHYGCGQVQGNL 60
DB 1 MELALCRWGLLALLPPGAASVQCTGTDMKRLPASPEHLDMLRLHYGCGQVQGNL 60
QY 61 ELYTPPTNASSFLQDIOENQGYVLIANNQVQVPLQRTVYRGQLFEDNYALAVLNG 120
DB 61 ELYTPPTNASSFLQDIOENQGYVLIANNQVQVPLQRTVYRGQLFEDNYALAVLNG 120
QY 121 DPLNNTTAVTGASPGGLRELDRLSTELILKGVLIQRNPQCYODTILMKDIFHNQOLA 180
DB 121 DPLNNTTAVTGASPGGLRELDRLSTELILKGVLIQRNPQCYODTILMKDIFHNQOLA 180
QY 181 LTLIDTNRSRACHPCSPKCKSRCKWGESSEDCQSLITRVCAAGCARCKGPLPTDCHEQC 240
DB 181 LTLIDTNRSRACHPCSPKCKSRCKWGESSEDCQSLITRVCAAGCARCKGPLPTDCHEQC 240
QY 241 AAGCTGPHNSOCLACLNHNSGICELHCPALVTYNTDTFESMPNDEGRYTGASCVTACP 300
DB 241 AAGCTGPHNSOCLACLNHNSGICELHCPALVTYNTDTFESMPNDEGRYTGASCVTACP 300
QY 301 YNYISTDVGSCTLVCPILNQEVTAEADGTQRCCKSKPCARVCYGI 345
DB 301 YNYISTDVGSCTLVCPILNQEVTAEADGTQRCCKSKPCARVTHSL 345

```

RESULT 9  
 ID Q8R2X1 PRELIMINARY; PRT; 367 AA.  
 AC Q8R2X1;  
 DT 01-JUN-2002 (Tremblrel. 21, Created)  
 DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Hypothetical 40.2 kDa protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.;  
 RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.



DR EMBL: BC027080; AA027080.1; -  
 KM Hypothetical protein.  
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;

Query Match 25.4%; Score 1739; DB 11; Length 367;  
 Best Local Similarity 88.0%; Pred. No. 1.3e-124;  
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;

QY 889 MALESILRRRTTQSDVWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 948  
 DB 1 MALESILRRRTTQSDVWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 60  
 QY 949 IDVYIMVKCMIMIDSECRPRFRELVESEFRMARDPQRFVIONEDLGPA5PLDSTFYRSL 1008  
 DB 61 IDVYIMVKCMIMIDSECRPRFRELVESEFRMARDPQRFVIONEDLGPA5PLDSTFYRSL 120  
 QY 1009 LEDDDMGDLVDAEEYLVPOQGFPCPDPAAGAGVWHRHRSSTRSGGDLTLGLPSEE 1068  
 DB 121 LEDDDMGDLVDAEEYLVPOQGFPCPDPAAGAGVWHRHRSSTRSGGDLTLGLPSEE 180  
 QY 1069 EAPRSPLARSEGAGSDVDFGDLGMAKAGLQSLPTHPSPPLQRYSEDPVPLPSETDGYV 1128  
 DB 181 EAPRSPLARSEGAGSDVDFGDLGMAKAGLQSLPTHPSPPLQRYSEDPVPLPSETDGYV 240  
 QY 1129 APLTCSPOPEYVNPQDVRPQPSREGPLPAARPAATLERAKTSLPGKGVVYKDVFAFG 1188  
 DB 241 APLTCSPOPEYVNPQDVRPQPSREGPLPAARPAATLERAKTSLPGKGVVYKDVFAFG 300  
 QY 1189 GAVENPEYLTPOGGAAPQPPAFSPAFDNLVYWDPPRCGAPSTFKGTPTAENPEY 1248  
 DB 301 GAVENPEYLTPOGGAAPQPPAFSPAFDNLVYWDPPRCGAPSTFKGTPTAENPEY 360  
 QY 1249 LGLDVFPV 1255  
 DB 361 LGLDVFPV 367

## RESULT 10

Q8WYV0 PRELIMINARY; PRT; 412 AA.

AC G8WYV0;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Hypothetical 44.7 kDa protein.  
 GN P33659.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,  
 RA Wan D.F., Gu J.R.;  
 RT "Novel human cDNA clones with function of inhibiting cancer cell  
 RT growth."  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF318349; AAL55856.1; -  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00219; TyKc; 1.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN 1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.  
 KM Hypothetical protein.  
 SQ SEQUENCE 412 AA; 44702 MW; 034397FF3F27D8C CRC64;

Query Match 24.8%; Score 1697.5; DB 4; Length 412;  
 Best Local Similarity 80.5%; Pred. No. 2.3e-121;

Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;

QY 889 MALESILRRRTTQSDVWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 948  
 DB 1 MALESILRRRTTQSDVWSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 60  
 QY 949 IDVYIMVKCMIMIDSECRPRFRELVESEFRMARDPQRFVIONEDLGPA5PLDSTFYRSL 1008  
 DB 61 IDVYIMVKCMIMIDSECRPRFRELVESEFRMARDPQRFVIONEDLGPA5PLDSTFYRSL 120  
 QY 1009 LEDDDMGDLVDAEEYLVPOQGFPCPDPAAGAGVWHRHRSSTRSGGDLTLGLPSEE 1068  
 DB 121 LEDDDMGDLVDAEEYLVPOQGFPCPDPAAGAGVWHRHRSSTRSGGDLTLGLPSEE 180  
 QY 1069 EAPRSPLARSEGAGSDVDFGDLGMAKAGLQSLPTHPSPPLQRYSEDPVPLPSETDGYV 1128  
 DB 181 EAPRSPLARSEGAGSDVDFGDLGMAKAGLQSLPTHPSPPLQRYSEDPVPLPSETDGYV 240  
 QY 1129 APLTCSPOPEYVNPQDVRPQPSREGPLPAARPAATLERAKTSLPGKGVVYKDVFAFG 1188  
 DB 241 APLTCSPOPEYVNPQDVRPQPSREGPLPAARPAATLERAKTSLPGKGVVYKDVFAFG 300  
 QY 1189 GAVENPEYLTPOGGAAPQPPAFSPAFDNLVYWDPPRCGAPSTFKGTPTAENPEY 1248  
 DB 301 GAVENPEYLTPOGGAAPQPPAFSPAFDNLVYWDPPRCGAPSTFKGTPTAENPEY 360  
 QY 1221 YWVD-QDPPER-----GAPSTFKGTPTAEN 1245  
 DB 361 YWVD-QDPPER-----GAPSTFKGTPTAEN 410

## RESULT 11

Q86712 PRELIMINARY; PRT; 729 AA.

AC Q86712;  
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)  
 DE Polypeptide.  
 GN POLYPEPTIDE.  
 OS Avian rous-associated virus type 1.  
 OC Viruses; Retroviridae; Retroviridae; Alpharetrovirus.  
 OX NCBI\_TaxID=11950;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94203659; PubMed=8152791;  
 RA Venustom B., Raynosccheck C., Jansson L., Doederlein G., Lhotak V.,  
 RA Johnson A., Beug H.;  
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
 RT different transforming capacities."  
 RL Oncogene 9:1307-1320(1994).  
 DR EMBL: S69372; AAC60725.1; -  
 DR HSP; P03322; 1A6S.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR004028; Retro\_M.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF02813; Retro\_M; 1.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00219; TyKc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KM ATP-binding; Transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 729 AA; 80649 MW; 84D2F914EFED63 CRC64;

Query Match 24.5%; Score 1675; DB 15; Length 729;  
 Best Local Similarity 54.0%; Pred. No. 2.8e-119;  
 Matches 352; Conservative 68; Mismatches 133; Indels 100; Gaps 13;

QY 569 PONGSVTCGEGADQVCACVAKYKPPFCVACRPSGVKXDLVYMPYKFPDEBAGACQPCPI 628

```

Db      141 PEETATPKTGP--DHCKMCAHFIDGHCVCACPAVLGENDTL-VKYNADANAACOLCHP 197
Qy      629 NCTHSCVDLDDKGCPCAPASPLTISFNNFTVSFWLRYVKVSAASHLEKROOKIRKYTMR 688
Db      198 NCTRGCKGPGLECCF---NGSKTPSIAGAVGGLCLVVGIGIGLYLRRIIVYKRLR 254
Qy      689 RLLOETELVEPLTPSGAMPNOQMRLKETELRKVKVLSGAFGVYKGIWPDGENVKI 748
Db      255 RLQERELVEPLTPSGEAPNOAHLRIKETEFKVKVLSGAFGVYKGIWPDGENVKI 314
Qy      749 PVAIKYLRRENTSPKANKELIDEAYVWAGVSPYVRLIGICTSTVQVLTOLMPYGCILD 808
Db      315 PVAIKELRENTSPKANKELIDEAYVWAGVSPYVRLIGICTSTVQVLTOLMPYGCILD 374
Qy      809 HVRENKRGRLGSDLLNMCQIAKMSYLEDVLRVLRDLAARNVLYKSPNHVKITDFGLAR 868
Db      375 YIREKMDNGSQVLLNMCQIAKMSYLEDVLRVLRDLAARNVLYKSPNHVKITDFGLAR 434
Qy      869 LLDIDETEHADGGKVPKIMMALESLIRRFTHQSDVWSYGVTVWELMTFGAKPYDGI 928
Db      435 LIGADEKEVHAEGKVPKIMMALESLIRRFTHQSDVWSYGVTVWELMTFGAKPYDGI 494
Qy      929 REIPDLKEXGERLPOPICTIDVYIMVCMIDSECRPRFRELVSFSPMARDPORFVY 968
Db      495 SEISSYLEKGERLPQPICTIDVYIMVCMIDSECRPRFRELVSFSPMARDPORFVY 554
Qy      989 IQ-NEDLGPASPLDSTFFYSLIEDDDMGDLVDAEYLYPQGFPCPDPAAGAGVHHRH 1047
Db      555 IQGDERMHLPSPTDSKFRYRLTMEEDMEDIVDAEYLYPQGFPCPDPAAGAGVHHRH 598
Qy      1048 RSSSTRSGGGDLTLGLEPSEEEAPRSPPL---APSEGGASDVFPDGLMGAKAGLQSLP 1102
Db      599 NSPST-----SRTPLLSSLSATSNSNATNCID-----RNGQCHP 632
Qy      1103 THDPSPLOQYSEDPTVPLPSET--DGVAAPLTCSPQPEYVNOQDVPAPSPREGPLPA 1160
Db      633 VREDSFVQYSSDPTGPNFLEESIDGFL-----PAPEYVNO--LMPKKPS----- 675
Qy      1161 RRAAGATLERAKTLSPCKNGVVDV-----AFGAVENPEYL 1197
Db      676 -----TAVYQNOIYNNISLTAISKLPMSRYONSHSTAVDNPEYL 715

RESULT 12
066714 PRELIMINARY: PRT; 567 AA.
ID 066714:
AC 086714:
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TrEMBLrel. 20, Last annotation update)
DE V-erbB protein (Fragment).
GN V-ERBB.
OS Avian rous-associated virus type 1.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCB1_Taxid=11950;
RN 11)
RP SEQUENCE FROM N.A.
RX MEDLINE=94203659; PubMed=8152791;
RA Vennstrom B., Raynosccheck C., Jansson L., Doederlein G., Lhotak V.,
RA Johnson A., Beug H.
RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid
RT evolution of distinct viral genomes carrying mutant v-erbB genes with
RT different transforming capacities."
RL Oncogene 9:1307-1320(1994).
DR EMBL; S69372; AAC60727.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR PRODOM; PD000001; Euk_pkinase; 1.
DR SMART; SMO0219; TYRC; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.

```

```

DR PROSITE; PS00107; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KW Tyrosine-protein kinase.
FT NON TER
SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7AD725E1 CRC64;

Query Match
Best Local Similarity 54.6%; Score 1673; DB 15; Length 567;
Matches 351; Conservative 66; Mismatches 126; Indels 100; Gaps 13;

Qy 578 GPEADOCVACAHYDPFCYACRCSSYKPLSLTMYPIKPFDEGACQPCPCINCHSCVDL 637
Db 1 GP--DHCKMCAHFIDGHCVCACPAVLGENDTL-VKYNADANAACOLCHPNCRGCKGP 57
Qy 638 DDKCPCAPASPLTISFNNFTVSFWLRYVKVSAASHLEKROOKIRKYTRRLQETELV 697
Db 58 GLECCF---NGSKTPSIAGAVGGLCLVVGIGIGLYLRRIIVYKRLRRLQERELV 114
Qy 698 EPLTPSGAMPNOQMRLKETELRKVKVLSGAFGVYKGIWPDGENVKI PVAIKYLR 757
Db 115 EPLTPSGEAPNOAHLRIKETEFKVKVLSGAFGVYKGIWPDGENVKI PVAIKELRE 174
Qy 758 NTSFKAKELIDEAYVWAGVSPYVRLIGICTSTVQVLTOLMPYGCILDHVENRGR 817
Db 175 ATSPKAKELIDEAYVWAGVSPYVRLIGICTSTVQVLTOLMPYGCILDYIREKMDN 234
Qy 818 GSQDLNMCQIAKMSYLEDVLRVLRDLAARNVLYKSPNHVKITDFGLARLLDIDETEV 877
Db 235 GSQVLLNMCQIAKMSYLEDVLRVLRDLAARNVLYKTPQHVKITDFGLAKLIGADEKE 294
Qy 878 HADGGKVPKIMMALESLIRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PAREIPDL 937
Db 295 HADGGKVPKIMMALESLIRRFTHQSDVWSYGVTVWELMTFGAKPYDGI PASEISYLEK 354
Qy 938 GERLPQPICTIDVYIMVCMIDSECRPRFRELVSFSPMARDPORFVYIQ-NEDLGP 996
Db 355 GERLPQPICTIDVYIMVCMIDSECRPRFRELVSFSPMARDPORFVYIQGDERMHL 414
Qy 997 ASPLDSTFFYSLIEDDDMGDLVDAEYLYPQGFPCPDPAAGAGVHHRHSSSTRSGG 1056
Db 415 PSPTDSKFRYRLTMEEDMEDIVDAEYLYPQGF-----NSPST----- 454
Qy 1057 GDULTLGLPESEEEAPRSPPL---APSEGGASDVFPDGLMGAKAGLQSLPTHPDPLQR 1111
Db 455 -----SRTPLLSSLSATSNSNATNCID-----RNGQCHPVREDSFVQR 492
Qy 1112 YSEDPTVPLPSET--DGVAAPLTCSPQPEYVNOQDVPAPSPREGPLPAPRAGATLER 1169
Db 493 YSSDPTGPNFLEESIDGFL-----PAPEYVNO--LMPKKPS----- 526
Qy 1170 AKTISPCKNGVVDV-----AFGAVENPEYL 1197
Db 527 -----TAVYQNOIYNNISLTAISKLPMSRYONSHSTAVDNPEYL 566

RESULT 13
064895 PRELIMINARY: PRT; 962 AA.
ID 064895:
AC 064895:
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Gag-V-erb-A-V-erb-B protein.
GN Gag-V-ERB-A-V-ERB-B.
OS Avian erythroblastosis virus.
OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.
OX NCB1_Taxid=11861;
RN 11)
RP SEQUENCE FROM N.A.
RX MEDLINE=90206603; PubMed=1369616;
RA Briskin A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;
RA "Six amino acids from the retroviral gene gag greatly enhance the
RT transforming potential of the oncogene v-erb-B";

```

CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 DR EMBL; X52209; CA36459.1; -  
 DR EMBL; X52211; CA36459.1; JOINED.  
 DR HSSP; P10828; ZNLF.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR000535; Hormone\_rec\_1lg.  
 DR InterPro; IPR001723; Stcdhmn\_receptor.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR InterPro; IPR001628; Znf\_C4steroid.  
 DR Pfam; PF00104; hormone\_rec; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00105; zfc4; 1.  
 DR PRINTS; PR00398; STRDHOMONER.  
 DR PRINTS; PR00047; STROIDFINGER.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR ProDom; PD000035; Znf\_C4steroid; 1.  
 DR SMART; SM00430; HOL1; 1.  
 DR SMART; SM00219; TYRKC; 1.  
 DR SMART; SM00399; Znf\_C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR ATP-binding; DNA-binding; Nuclear protein; Receptor;  
 KM Transcription regulation; Tyrosine-protein kinase;  
 KM Zinc-finger.  
 SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4B95CE CRC64;

Query Match 23.8%; Score 1608.5; DB 15; Length 962;  
 Best Local Similarity 50.9%; Pred. No. 5.2e-114;  
 Matches 352; Conservative 66; Mismatches 156; Indels 117; Gaps 16;

QY 541 VEECRVLQGPREF-YNNAR-HCLP-----CHPEQC 568  
 DB 354 IEKQESYLLAFHYNYRGNIPHWKILMKVADLMIGAYHARFLMKVECPTELS 413  
 QY 569 PONGSVTCGPEADQCVACHYDPPFCVACPSGVKPDLSYMPKPFDEGACQPCPI 628  
 DB 414 PQE-----VGP--DHCKKCAHFIDGPHCVKACAPAGVGENDTL-VKRYADANAVCQLCHP 465  
 QY 629 NCHSCVDDDKCKCPAQRASPLTSTFNNFTVSFWLRVPKVSASHLEKROOKIRKYTKR 688  
 DB 466 NCRGCKGPGLECCP--NGSKTPSIAGVVGGLCLVAVGGLGILYLRHRHIVRRKTLR 522  
 QY 689 RLQETELVEPLTPSGAMPNOAQMRILKETELRKVKVLSGAFGTYKXIMIPDENVKI 748  
 DB 523 RLQGERELVEPLTPSGEARPOAHRLIKETEFKVKVLPVGGAGVYKGLMIPBEGKYTI 582  
 QY 749 PVAIKVLRENTSFKANKELIDEAYVYAGVSPYVSRLLGICLTSTVQLVTQLMPYGCILLD 808  
 DB 583 PVAIKELRENTSFKANKELIDEAYVYASVDNPHVCRLLGICLTSTVQLVTQLMPYGCILLD 642  
 QY 809 HYRENKGRIGSODLLMWCMQIAKMSYLEDVLRHNDLAARVLYKSPNHVKITDPGLAR 868  
 DB 643 YIREHNDIGSQYLLMWCVQIAKGMVLEERHNVHNDLAARVLYKTPQHVKITDPGLAK 702  
 QY 869 LLDIDETEVHADGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPA 928  
 DB 703 QCGADEKEYHAEGKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPA 762  
 QY 929 RLIPDLLEKGERLPQPICTIDVYIMVYKCMIDSECRPFRELVESEFARMADPQRFVY 988  
 DB 763 SEISSVLEKGERLPQPICTIDVYIMVYKCMMSGADSRKPFRELLIAEFKMARDPRIYLV 822  
 QY 989 IO-NEDLAGPASELSTFFYSLLLEDDMDGLVDAEELVYVQGFQCPDPAFGAGVNHHRH 1047  
 DB 823 IGDGERMHLPSPTDSKFFYRTIMEEEMEDIVDAEYLDPHQGF----- 866  
 QY 1048 RRSSTRSGGGDLTLGLPESEEAAPSPLAPSEGAGSDVFDGLGAKAGLQSLTPHDPS 1107  
 DB 867 NSPST-----SRTPLLSLSATSN-----NSATKCIDRNGH--- 898

QY 1108 PLQRYSEDPYVLPSEITDGYVAPLTCSPQPEYVNPQVYRPPQPSREGBLPAPAPAGAT- 1166  
 DB 899 -----PVREDOFL-----PAPEVYVQ--LMPKPSSTAMVQNYNYSILTA 937

QY 1167 LERATLSPGKGVKVDVPAFGAVENREYL 1197  
 DB 938 ISKLPMDSRYQN-----SHSTAVDNREYL 961

RESULT 14  
 Q85468  
 ID Q85468 PRELIMINARY; PRT; 545 AA.  
 AC Q85468;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Avian erythroblastosis virus (T834) v-erbB gene.  
 OS Avian erythroblastosis virus.  
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.  
 NX NCB\_Taxid=11861;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88217326; PubMed=2897102;  
 RA Scotting P., Vennstrom B., Jansen M., Graf T., Beug H., Haymann M.J.;  
 RT "Common site of mutation in the erbB gene of avian erythroblastosis  
 RT virus mutants that are temperature sensitive for transformation."  
 RL Oncogene Res. 1:265-278(1987).  
 DR EMBL; X06943; CA36024.1; -  
 DR HSSP; P11362; IFGK.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00219; TYRKC; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KM ATP-binding; Tyrosine-protein kinase.  
 SQ SEQUENCE 545 AA; 60899 MW; 140DC8C8CA0F8AF4 CRC64;

Query Match 23.4%; Score 1600; DB 15; Length 545;  
 Best Local Similarity 54.1%; Pred. No. 9.9e-114;  
 Matches 339; Conservative 63; Mismatches 135; Indels 90; Gaps 13;

QY 578 GPEADQCVACHYDPPFCVACPSGVKPDLSYMPKPFDEGACQPCPINCHSCVDL 637  
 DB 1 GP--DHCKKCAHFIDGPHCVKACAPAGVGENDTL-VKRYADANAVCQLCHPNCRCGCP 57  
 QY 638 DQKGPAPQARASPLTSTFNNFTVSFWLRVPKVSASHLEKROOKIRKYTMRELLOETELV 697  
 DB 58 GLEGGP--NGSKTPSIAGVVGGLCLVAVGGLGILYLRHRHIVRRKTLRLOERELV 114  
 QY 698 EPLTPSGAMPNOAQMRILKETELRKVKVLSGAFGTYKXIMIPDENVKPVAIKVLE 757  
 DB 115 EPLTPSGEARPOAHRLIKETEFKVKVLPVGGAGVYKGLMIPBEGKYTI PVAIKELRE 174  
 QY 758 NTSRKANKELIDEAYVYAGVSPYVSRLLGICLTSTVQLVTQLMPYGCILLDHYENRRL 817  
 DB 175 ATSPRANKELIDEAYVYASVDNPHVCRLLGICLTSTVQLVTQLMPYGCILLDYIREHNDI 234  
 QY 818 GSODLLMWCMQIAKMSYLEDVLRHNDLAARVLYKSPNHVKITDPGLARLLDIDENEY 877  
 DB 235 GSQYLLMWCVQIAKGMVLEERHNVHNDLAARVLYKTPQHVKITDPGLAKQGADEKEY 294  
 QY 878 HADGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPARBIPLLEK 937  
 DB 295 HAEQKVPKIMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASEISSVLEK 354  
 QY 938 GERLPQPICTIDVYIMVYKCMIDSECRPFRELVESEFARMADPQRFVYVQ-NEDLGP 996  
 DB 355 GERLPQPICTIDVYIMVYKCMMSDADSRKPFRELLIAEFKMARDPRIYLVIGQDERMHL 414

QY 997 ASPLDSTFYRLSDDDMDGLVDAEETLVPQGFPCPPAPAGAGVYHHRSSSTRSG 1056  
 DB 415 PSPTDSKFRYTLMEEDMEDIVADELYVHQGF-----NSPT-----454  
 QY 1057 GDILTLGLEPSEEPSPPL-----APREGASVFPQDGLGMAKGLQSLPHYDPSPLOR 1111  
 DB 455 -----SRTPILSSLSATSNNSATNCIPRNG-----H-----481  
 QY 1112 YSEDPYVLPSETDGYVAPLTCSPQPEYVQDPVDPQPSREGPLPAAPAGAT-LEBA 1170  
 DB 482 -----PVREDGF-----PAREYVQ--LMKKSTANVQIQINYSILTAISKL 524  
 QY 1171 KTLSPKNGVVKDVFAPGAVENPEYL 1197  
 DB 525 PMDSRYQN-----SHSTAVDNPEYL 544

## RESULT 15

Q9WVF5 PRELIMINARY; PRT: 655 AA.  
 ID Q9WVF5  
 AC Q9WVF5  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).  
 GN EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID:10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=LIVER;  
 RA Reiter J.L., Thredgill D.W., Danielsen A.J., Schnell C., Lampand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maibhe N.J.;  
 RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-terminal truncated Receptors."  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CH/101, 129/SVJ, AND 129/SVEVTC;  
 RA Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielsen A.J., Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampand A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maibhe N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=LIVER;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Akawa T., Hata A., Fukunishi Y., Konno H., Adachi N., Fukuoka K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kiehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Balderrelli R., Barsch G., Blake J., Boffelli D., Bojunga N., Carinici P., de Bonaldo M.F., Brownstein M.J., Bull C., Fletcher C., Fujita K., Garioldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaeerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S., Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection."  
 RL Nature 409:685-690(2001).  
 DR EMBL, AF124513; AA044149.1; -  
 DR EMBL, AF275365; AAG28047.1; -  
 DR EMBL, AF275364; AAG28047.1; JOINED.  
 DR EMBL, AF275365; AAG28047.1; JOINED.  
 DR EMBL, AK004944; BAB23688.1; -  
 DR EMBL, AK004883; BAB23641.1; -  
 DR EMBL, AK004911; BAB23662.1; -  
 DR MGD, MGI:95294; Egfr.  
 DR InterPro, IPR00494; EGFR\_L domain.  
 DR InterPro, IPR002174; Furin-like.  
 DR Pfam, PF00757; Furin-like.  
 DR Pfam, PF01030; Recep\_L\_domain; 2.  
 DR SMART, SMO0261; FU; 3.  
 KW Receptor.  
 SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.0%; Score 1506.5; DB 11; Length 655;  
 Best Local Similarity 44.5%; Pred. No. 1.8e-106;  
 Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

QY 11 LTLALPPGAA--STQVCTGDMKRLPASETHLDMLHYOGGVQVGLTYLPTN 68  
 DB 14 LTLALCAAGALIEKKVCGTNSRLTQJGTFEDHLSLQRYNCEVVLGNLEITYVORN 73  
 QY 69 ASLSFLQDIOEVGYVLLIAHNVQVPLQRLRIYRGTLQFEDNYALAVLNDGDPPLNTP 128  
 DB 74 YDLSTLKTIOEYAGVLLALNTVERIPLENQIRGNALYNTALATLS----- 124  
 QY 129 VTGASPGRLREQLRLSTELLKGVLLIQNPQLCYQDTILMKD---FHQNOALATLI 184  
 DB 125 -YGTNRTGLRELPMNMLEILIGAVFSNPNILCNMDIQRDVIQVNFMSMSMDL--- 180  
 QY 185 DTRNRACHPSPCMKSGRCGSESDQSLRTVCAGCA-RCKGPLPTDCHQCAAG 243  
 DB 181 -QSHSSCPKCDSPNSCWGGEBENCKLTKITCAQCCCHRCGRSPDCHQCAAG 239  
 QY 244 CTGPHSDCLACLHFNHSGELCHCPALVTYNTDTFESMPNPEGRTFGASCVTACPNY 303  
 DB 240 CTGPRSPDCLVQKFDATCKDTCPPMLNPTTYQMDVNPCKYSGATCVKCKCPNY 299  
 QY 304 LSTVSGCTVCPRLNCEVTADGTQREKSKCARCYGLQMYIANKSFTIGITELE 363  
 DB 300 VTDHGSCTVACGPDYEV-EDGIRKCKCDGPKKVCNIGIGEEK-DTLSTINATIK 357  
 QY 364 -FAGCKITFGSLAFIPESFGDPPASNTAPLQPEQLOVETLEETITGYLISAMPDLPDL 422  
 DB 358 HFXYTALSGDLHLFVAFKXDSFRTPLDPRLEILIKYKEITGFLIQAMPDNWTDL 417  
 QY 423 SVFQNLQYIRGRLHNGVSLTLOGLSIWLGLSLRLSGLALIHNTLCFVHTVPM 482  
 DB 418 HAFENLEIRGRTRKHGQFSLAVVGLNTISGLSLKELISGDVYIISGNRLCYANTINW 477  
 QY 483 DOLFENPQALHTANRPEDECVSEGACLOLARGHWGGPFGTCVNCQFLLGOEVE 542  
 DB 478 KKLFTPTQKTKIMNRKCKOKAVNHCNLCSEGGWGPEDDVCSCVNSGREVE 537  
 QY 538 KCNILEGPRFEVENSECICQHPCLPQANNITCTGRPDNCICAHYIDGPHCVKTCFA 597  
 DB 543 ECRVLQGLPREYVNAHGLCPHPCOPONGSVTCFEPADQVCAAHYKDPFVCARCS 602  
 QY 603 GVKPDLSTYMPKPPDEGACQPCPINTHSCVLDLDDGC 642  
 DB 598 GIMGENNTL-VKRTADANNVCHLGHANTCYCAGGGLQGC 636

Search completed: July 22, 2003, 09:24:19  
 Job time : 51.8246 secs

Result No.	Score	Query Match	Length	DB	ID	Description
1	6627	97.0	1255	21	AA932620	Human heregulin 1
2	6627	97.0	1255	22	AAE12130	Human tyrosine kinase
3	6627	97.0	1255	22	AAE60145	HER2 transgene plasmid
4	6627	97.0	1255	23	AAU74645	Human HER2 (ErbB2)
5	6623	96.9	1255	17	AAW01111	Human HER2 protein
6	6623	96.9	1255	20	AAW92406	Human HER-2/neu protein
7	6623	96.9	1255	21	AAAB21198	Human HER-2/neu protein
8	6623	96.9	1255	21	AAV84780	Human acid sequence
9	6623	96.9	1255	22	AAAB5458	Human HER-2/neu protein
10	6623	96.9	1255	22	AAAG8267	HER2/neu amino acid

[illegible]

FT /label= Cysteine\_rich\_domain  
 FT Region  
 FT 210..224  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 250..264  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 324..483  
 FT /label= Ligand\_binding\_domain  
 FT 325..339  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 369..383  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 465..479  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 484..623  
 FT /label= Cysteine\_rich\_domain  
 FT 579..593  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 624..654  
 FT /label= Transmembrane\_domain  
 FT 632..652  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 653..667  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 655..1010  
 FT /label= Tyrosine\_kinase\_domain  
 FT 661..675  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 695..709  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 710..730  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 1011..1235  
 FT /label= C-terminal\_domain  
 PN MO200020027-A2.  
 PD 13-APR-2000.  
 XX 05-OCT-1999; 99WO-DK00525.  
 XX 05-OCT-1998; 98DK-0001261.  
 PR 20-OCT-1998; 98US-0105011.  
 XX (MEBT-) M & E BIOTECH AS.  
 PA Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Ieash D, Dalum I;  
 PI Gautam A, Birk P, Karlsson G;  
 XX WPI: 2000-349917/30.  
 DR N-ESDB: AAA09455.  
 XX Inducing immune responses to weakly immunogenic, tumor associated  
 PT peptide antigens for the treatment of breast and prostate cancer  
 XX  
 PS Claim 62; Page 193-198; 220pp; English.  
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of  
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL  
 CC response. Subdominant CTL epitopes, antibody binding regions and  
 CC cysteine residues involved in disulfide bonds are preserved in the  
 CC immunogenized forms. Regions suitable for the insertion of foreign T  
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic  
 CC cell-associated peptide antigens (PA) such as those associated with  
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen  
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).  
 CC The method comprises effecting simultaneous presentation by antigen  
 CC producing cells (APCs) of the animal's immune system of: (1) at least 1  
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1  
 CC B-cell group derived from the cell-associated PA; and (2) at least 1  
 CC first T helper cell group which is foreign to the animal. Analogues of  
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial  
 CC part of all known and predicted CTL and B-cell epitopes of the respective  
 CC PA and including at least one foreign T helper epitope are also claimed.  
 CC The method is used to treat prostate, prostate/breast or breast cancer  
 CC when the PA is human PSM, FGF8b and Her2, respectively.  
 CC  
 XX  
 SO Sequence 1255 AA;

Query Match 97.0%; Score 6627; DB 21; Length 1255;  
 Best Local Similarity 97.4%; Pred. No. 0;  
 Matches 1222; Conservative 6; Mismatches 27; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALLPFAASTQVCTGDMRLRPASPETHLMDLRLHYQCCQVQGNL 60  
 Db 1 MELAALCRWGLLALLPFAASTQVCTGDMRLRPASPETHLMDLRLHYQCCQVQGNL 60  
 QY 61 ELTYLPTNASLSPLODIQEVQGYVLIANQVQVPLQRLIRYGTQLFEDNYALAVLNG 120  
 Db 61 ELTYLPTNASLSPLODIQEVQGYVLIANQVQVPLQRLIRYGTQLFEDNYALAVLNG 120  
 QY 121 DPANNTPTVGASPGGLRELOLRSLTEILKGVLLQBNQOLQYOTILMKDIFHNQOLA 180  
 Db 121 DPANNTPTVGASPGGLRELOLRSLTEILKGVLLQBNQOLQYOTILMKDIFHNQOLA 180  
 QY 181 LTLIDNRSRACHPSPCKSGRCWSESESDQSLTRVYACAGCARCKPLPTDCHEQC 240  
 Db 181 LTLIDNRSRACHPSPCKSGRCWSESESDQSLTRVYACAGCARCKPLPTDCHEQC 240  
 QY 241 AAGCTGPKHSDCLACHFPHSGICELHCAALTYNTDTPFESPNREGRTFASCTYACP 300  
 Db 241 AAGCTGPKHSDCLACHFPHSGICELHCAALTYNTDTPFESPNREGRTFASCTYACP 300  
 QY 301 YVYLSTDVSGCTLVCPFLHQEVTAEQDQRCCKSPCARVCYGGMCQYIKANSKFIGIT 360  
 Db 301 YVYLSTDVSGCTLVCPFLHQEVTAEQDQRCCKSPCARVCYGGMCQYIKANSKFIGIT 360  
 QY 361 ELDFAGCKKIFGSLAFLPSPGDPASNTAPLPQPOLQVFTLEETIGLYISAMPDSLIP 420  
 Db 361 IOEFAGCKKIFGSLAFLPSPGDPASNTAPLPQPOLQVFTLEETIGLYISAMPDSLIP 420  
 QY 421 DLSVFQNLQVIRGRIHNGAYSLTQIGISWLGRLSRLRELSGGLALIHNNHLQFVHTV 480  
 Db 421 DLSVFQNLQVIRGRIHNGAYSLTQIGISWLGRLSRLRELSGGLALIHNNHLQFVHTV 480  
 QY 481 PMDQLFRNHQALLHTANRPEDECVGEGLAQHQLCARHGWPGPTQVCNCSQFLRGQEC 540  
 Db 481 PMDQLFRNHQALLHTANRPEDECVGEGLAQHQLCARHGWPGPTQVCNCSQFLRGQEC 540  
 QY 541 VEECRVLQGLPREYVNAHCLPCHPECCPQNSVTCFGEADQCVACHYKDPFCVARC 600  
 Db 541 VEECRVLQGLPREYVNAHCLPCHPECCPQNSVTCFGEADQCVACHYKDPFCVARC 600  
 QY 601 PSGVXPDLSTYMPDIWPFDEGACQCPINCTHSCVDLDKGPAPQARASPLTSINNFNTV 660  
 Db 601 PSGVXPDLSTYMPDIWPFDEGACQCPINCTHSCVDLDKGPAPQARASPLTSINNFNTV 660  
 QY 661 SFWLAVPKYSASHLEKRRQKTRKTYMRLLQETLVLEPLTPSGAMPQAOQRILKETEL 720  
 Db 661 ILVVAVLGVFGIILKRRQKTRKTYMRLLQETLVLEPLTPSGAMPQAOQRILKETEL 720  
 QY 721 RKVKYLGSGAFYTVKGIWIPDGEVVKIPVAIKVLRNTSPRANKIIDEAVYVAGVSP 780  
 Db 721 RKVKYLGSGAFYTVKGIWIPDGEVVKIPVAIKVLRNTSPRANKIIDEAVYVAGVSP 780

QY 781 YVSRLLGICLTSTVOLVLTQMLPFGCLLDHVRNRRGLSGSODLLNMCMQIAKMSYLEDVR 840  
 DB 781 YVSRLLGICLTSTVOLVLTQMLPFGCLLDHVRNRRGLSGSODLLNMCMQIAKMSYLEDVR 840  
 QY 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLIDETETVHADGGKVP:IKMMALESILRRRT 900  
 DB 841 LVHRDLAARNVLYKSPNHVKITDFGLARLLIDETETVHADGGKVP:IKMMALESILRRRT 900  
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDG:IPAREIPDLLEKGERLPQPICTIDVYMIWKCWM 960  
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDG:IPAREIPDLLEKGERLPQPICTIDVYMIWKCWM 960  
 QY 961 IDSECRPRFRELSESRMARDDPREFVIONEDLGPAASPLDSFFVSLLEDDMDGLVA 1020  
 DB 961 IDSECRPRFRELSESRMARDDPREFVIONEDLGPAASPLDSFFVSLLEDDMDGLVA 1020  
 QY 1021 EBYLVDPQGFCDPAPAGAGVNHRRSSSTRSGGDDTLGLEPSEEEAPRSPPLAFSEG 1080  
 DB 1021 EBYLVDPQGFCDPAPAGAGVNHRRSSSTRSGGDDTLGLEPSEEEAPRSPPLAFSEG 1080  
 QY 1081 AGSDVDFDGLGMAAAGLOSLPHTDPSPLQRYSEDPTVPLPSTDDYVAPLTCSPQPEYV 1140  
 DB 1081 AGSDVDFDGLGMAAAGLOSLPHTDPSPLQRYSEDPTVPLPSTDDYVAPLTCSPQPEYV 1140  
 QY 1141 NQPDVBPQPSPEEGPLPAARPAAGATLEBAKTLSPQKNGVVDVFAFGAVENPEYLTPQ 1200  
 DB 1141 NQPDVBPQPSPEEGPLPAARPAAGATLEBAKTLSPQKNGVVDVFAFGAVENPEYLTPQ 1200  
 QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQPPERGAPESTFKGPTAENPEYGLDVPV 1255  
 DB 1201 GGAAPQHPPPAPSPAFDNLVYWDQPPERGAPESTFKGPTAENPEYGLDVPV 1255

RESULT 2  
 AAEL12130  
 ID AAEL12130 standard; Protein; 1255 AA.

XX AAEL12130;  
 AC 18-DEC-2001 (first entry)  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Human tyrosine kinase-type receptor, HER-2.  
 KW Therapeutic compound; major histocompatibility complex; vaccine;  
 KW anti-tumor peptide; MHC; immunoregulatory; immune response; HER-2;  
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;  
 KW antigen presenting cell; human; tyrosine kinase-type receptor.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 774..782  
 FT /note="Antigenic epitope"  
 XX  
 XX WO200168677-A2.  
 PD 20-SEP-2001.  
 XX  
 PF 16-MAR-2001; 2001WO-US40328.  
 XX  
 PR 16-MAR-2000; 2000US-0527487.  
 XX  
 PA (GENZ ) GENZYME CORP.  
 XX  
 PI Nicolette CA;  
 XX  
 XX WPI; 2001-616284/71.  
 DR N-PSDB; AAD19731.  
 XX  
 PT Novel synthetic therapeutic compound for inducing immune response and  
 PT for use in adoptive immunotherapy, has enhanced binding to major  
 PT histocompatibility molecules and enhanced immunoregulatory properties

XX Claim 4; Page 63-67; 69pp; English.

CC The invention relates to synthetic therapeutic compounds (antigenic  
 CC peptides) with enhanced binding to major histocompatibility complex  
 CC (MHC) molecules and enhanced immunoregulatory properties relative  
 CC to their natural counterparts. Compounds of the invention are useful  
 CC for inducing an immune response in a subject and for use in adoptive  
 CC immunotherapy. They are useful as components of anti-cancer vaccines  
 CC and to expand immune effector cells that are specific for cancers  
 CC characterised by expression of the breast cancer antigen, HER-2.  
 CC Polynucleotides that encode peptides of the invention are useful as  
 CC hybridisation probes and as primers for the detection of genes of gene  
 CC transcripts that are expressed in antigen presenting cells (APCs), to  
 CC confirm transduction of polynucleotides into host cells. The present  
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds  
 CC of the invention are designed based on the HER-2 antigenic peptide  
 CC (774-782).

SQ Sequence 1255 AA;

Query Match 97.0%; Score 6627; DB 22; Length 1255;  
 Best Local Similarity 97.4%; Pred. No. 0;  
 Matches 1222; Conservative 6; Mismatches 27; Indels 0; Gaps 0;

QY 1 MELALCGWGLLLALPPGASTOVCTGDKMLRPLASPEHMLMLRLYGGCQVQGNL 60  
 DB 1 MELALCGWGLLLALPPGASTOVCTGDKMLRPLASPEHMLMLRLYGGCQVQGNL 60  
 QY 61 ELTYLPTNASLSFLODIOEVQGYVLIANQVQVPLQRLIRVGTQLPEDNYALAVLDNG 120  
 DB 61 ELTYLPTNASLSFLODIOEVQGYVLIANQVQVPLQRLIRVGTQLPEDNYALAVLDNG 120  
 QY 121 DPLNNTPTVYASFGELRELOLRSLTEILKGVLIQNNPOLCYODTILMKDIFHNQOLA 180  
 DB 121 DPLNNTPTVYASFGELRELOLRSLTEILKGVLIQNNPOLCYODTILMKDIFHNQOLA 180  
 QY 121 DPLNNTPTVYASFGELRELOLRSLTEILKGVLIQNNPOLCYODTILMKDIFHNQOLA 180  
 DB 121 DPLNNTPTVYASFGELRELOLRSLTEILKGVLIQNNPOLCYODTILMKDIFHNQOLA 180  
 QY 181 LTLIDNRSRACHPCSPKCSRCMGSSSEDCSLRTVCAAGCARCGPLPTCCHEQC 240  
 DB 181 LTLIDNRSRACHPCSPKCSRCMGSSSEDCSLRTVCAAGCARCGPLPTCCHEQC 240  
 QY 241 AAGCTGPRHSDCLACLFHNSGICELCPALVTYNTDFFESMPNPEGVYTGACVYACP 300  
 DB 241 AAGCTGPRHSDCLACLFHNSGICELCPALVTYNTDFFESMPNPEGVYTGACVYACP 300  
 QY 301 YNYISTDVGSCTIVCPILNNOVTATDGTQREKSKRCARVYGLGMOYIRANKKFTGIT 360  
 DB 301 YNYISTDVGSCTIVCPILNNOVTATDGTQREKSKRCARVYGLGMOYIRANKKFTGIT 360  
 QY 301 YNYISTDVGSCTIVCPILNNOVTATDGTQREKSKRCARVYGLGMOYIRANKKFTGIT 360  
 DB 301 YNYISTDVGSCTIVCPILNNOVTATDGTQREKSKRCARVYGLGMOYIRANKKFTGIT 360  
 QY 361 ELEFAGCKKIFGSLAFLEPESFDGDPASNTAPLOPEOLQVFETLEITGYLISAMPDLP 420  
 DB 361 IQEFAGCKKIFGSLAFLEPESFDGDPASNTAPLOPEOLQVFETLEITGYLISAMPDLP 420  
 QY 421 DLSVFNQILQVIRGILHNQAVSLTLQGLISWLGRLRLREISGLALIHNTLCEVHTV 480  
 DB 421 DLSVFNQILQVIRGILHNQAVSLTLQGLISWLGRLRLREISGLALIHNTLCEVHTV 480  
 QY 481 PMDQLFNPHALHTANRPEDECVGSLACHOLCAAGHCGKGPPTCCVNCOSQRLQEC 540  
 DB 481 PMDQLFNPHALHTANRPEDECVGSLACHOLCAAGHCGKGPPTCCVNCOSQRLQEC 540  
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCGPEADQCVACAHYKDPFCVAC 600  
 DB 541 VEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCGPEADQCVACAHYKDPFCVAC 600  
 QY 601 PSQVPRDLSYMPITKKEFDEEGACQPCPINCTHSCVDLDDKGCPEAGASPLTSIVAVG 660  
 DB 601 PSQVPRDLSYMPITKKEFDEEGACQPCPINCTHSCVDLDDKGCPEAGASPLTSIVAVG 660  
 QY 661 SFMLRVPKVASHLERQOKIRKYTRRLQETELVEPLTPSGAMPNOMORILKETEL 720  
 DB 661 ILVVVLGVVFGILIKERQOKIRKYTRRLQETELVEPLTPSGAMPNOMORILKETEL 720

QY 721 RRYKVLGSGAFGTYYKGIWI PDGENVKI PVAIKVLRRENTSPKANKELIDEAYMAGVGS 780  
 DB 721 RRYKVLGSGAFGTYYKGIWI PDGENVKI PVAIKVLRRENTSPKANKELIDEAYMAGVGS 780  
 QY 781 YVSRLLGICLTSTVQVLTQMLPVGCLLDHVRNENRGLSGODLNNCMQIAKMSYLEDR 840  
 DB 781 YVSRLLGICLTSTVQVLTQMLPVGCLLDHVRNENRGLSGODLNNCMQIAKMSYLEDR 840  
 QY 841 LVHRDLAARNVLYVSPNHVKITDPGLARLLIDETETHAAGKVPIMKMALESILRRFT 900  
 DB 841 LVHRDLAARNVLYVSPNHVKITDPGLARLLIDETETHAAGKVPIMKMALESILRRFT 900  
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPCTTIDVYIMYKCM 960  
 DB 901 HQSDVMSYGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPCTTIDVYIMYKCM 960  
 QY 961 IDSECRPRFRELVEFSRMAADPQRFVVIQNEDEGPASPLDSTFYRSLDEDDMDGLVDA 1020  
 DB 961 IDSECRPRFRELVEFSRMAADPQRFVVIQNEDEGPASPLDSTFYRSLDEDDMDGLVDA 1020  
 QY 1021 EBYLVPOQGFECPPAPAGAGMVAHRRSSSTRSGGDLTLGLEPSEEA PRSPLAPSEG 1080  
 DB 1021 EBYLVPOQGFECPPAPAGAGMVAHRRSSSTRSGGDLTLGLEPSEEA PRSPLAPSEG 1080  
 QY 1081 AGSDVFPDGLCMGAKGLQSLPTHDSPLOQRYSEDPVPLPSETDGVAPLTGSPQPEYV 1140  
 DB 1081 AGSDVFPDGLCMGAKGLQSLPTHDSPLOQRYSEDPVPLPSETDGVAPLTGSPQPEYV 1140  
 QY 1141 NOPVRPQPSRPGSPPLAARPAAGATLERAKTSLPGKNGVYKDVFAFGAVENPEYLTPO 1200  
 DB 1141 NOPVRPQPSRPGSPPLAARPAAGATLERAKTSLPGKNGVYKDVFAFGAVENPEYLTPO 1200  
 QY 1201 GGAAPQHPPEPSPAFDNLVYWDQPPERGA PPTSTKGPPLAENPEYLDLDPV 1255  
 DB 1201 GGAAPQHPPEPSPAFDNLVYWDQPPERGA PPTSTKGPPLAENPEYLDLDPV 1255  
 QY 1255 GGAAPQHPPEPSPAFDNLVYWDQPPERGA PPTSTKGPPLAENPEYLDLDPV 1255  
 DB 1255 GGAAPQHPPEPSPAFDNLVYWDQPPERGA PPTSTKGPPLAENPEYLDLDPV 1255

RESULT 3  
 AAB60167  
 ID AAB60167 standard; Protein; 1255 AA.  
 AC AAB60167;  
 XX  
 DT 03-APR-2001 (first entry)  
 DE HER2 transgene plasmid construct encoded protein.  
 XX  
 KW Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;  
 XX antibody.  
 OS Homo sapiens.  
 XX  
 OS Synthetic.  
 XX  
 FN W0200100244-A2.  
 XX  
 PD 04-JAN-2001.  
 PF 23-JUN-2000; 2000WO-US17229.  
 XX  
 ER 25-JUN-1999; 99US-0141316.  
 PR 16-MAR-2000; 2000US-0189844.  
 XX  
 PA (GETH ) GENENTECH INC.  
 XX  
 PI Erickson S, Schwall R;  
 XX  
 DR WPI; 2001-061962/07.  
 DR N-PSDB; AAF24297.  
 XX  
 PT Treating tumors, particularly breast cancers, which overexpress an ErbB  
 XX receptor and does not respond to an anti-ErbB antibody, comprises  
 XX conjugating the antibody to a maytansinoid -

PS Example 3; Fig 4; 92pp; English.  
 XX  
 CC The present invention provides a method of treating cancer by  
 CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In  
 CC particular, the antibody is directed against ErbB2 (also known as HER2  
 CC and p185neu). The method is particularly useful in the treatment of  
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,  
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.  
 XX  
 SO Sequence 1255 AA;  
 Query Match 97.0%; Score 6627; DB 22; Length 1255;  
 Best Local Similarity 97.4%; Pred. No. 0;  
 Matches 1222; Conservative 6; Mismatches 27; Indels 0; Gaps 0;  
 QY 1 MELALCRWGLLALLPFGAASGYCTGDMKLRLPAPETHLDMRLHYGCGVVGNTL 60  
 DB 1 MELALCRWGLLALLPFGAASGYCTGDMKLRLPAPETHLDMRLHYGCGVVGNTL 60  
 QY 61 ELTYLPTNASLSFLQDIEVGQYVLIANOVQVELORLIRVGTQLFEDNYALAVDNG 120  
 DB 61 ELTYLPTNASLSFLQDIEVGQYVLIANOVQVELORLIRVGTQLFEDNYALAVDNG 120  
 QY 121 DPLNNTPTTASPGGLRELQRLSLTLIKGVLIQRNPOLCYOFTILMKDIFHKNNOLA 180  
 DB 121 DPLNNTPTTASPGGLRELQRLSLTLIKGVLIQRNPOLCYOFTILMKDIFHKNNOLA 180  
 QY 181 LTLIDITNRSRACHPSPCKSGSRGSESESDCSLTRVYACAGCARCKGPLPTDCHEQC 240  
 DB 181 LTLIDITNRSRACHPSPCKSGSRGSESESDCSLTRVYACAGCARCKGPLPTDCHEQC 240  
 QY 241 AAGCTGPKASDCLACHFNHSGICELHCPALVYNTDTFESMPNEBGRYTFGASCYTACP 300  
 DB 241 AAGCTGPKASDCLACHFNHSGICELHCPALVYNTDTFESMPNEBGRYTFGASCYTACP 300  
 QY 301 YNYLSTDVSGCTLVCPBLHNOEYTAEDGQRCCKSPCARVYCYGMOYIKANSKFI 360  
 DB 301 YNYLSTDVSGCTLVCPBLHNOEYTAEDGQRCCKSPCARVYCYGMOYIKANSKFI 360  
 QY 361 ELEFAGCKKI FGSFLAFLEPSFDGDPASNTAPLOPEOLOVFEETLEITGYLYISAMPDSL 420  
 DB 361 ELEFAGCKKI FGSFLAFLEPSFDGDPASNTAPLOPEOLOVFEETLEITGYLYISAMPDSL 420  
 QY 421 DLSVFQNLQVIGRILHNGAYSLTQSGISWLGSRSLRELGSSGLALLHNNHLCFVRHV 480  
 DB 421 DLSVFQNLQVIGRILHNGAYSLTQSGISWLGSRSLRELGSSGLALLHNNHLCFVRHV 480  
 QY 481 PMDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGCMGPGPTQCVNCSQFLRGQEC 540  
 DB 481 PMDQLFRNPHQALLHTANRPEDECVGEGLACHQLCARGCMGPGPTQCVNCSQFLRGQEC 540  
 QY 541 VEECRVLOGLPREYVNAHGLPCHEBCOPONGSVTCFGEADQCVACAHYKDPFVCVARC 600  
 DB 541 VEECRVLOGLPREYVNAHGLPCHEBCOPONGSVTCFGEADQCVACAHYKDPFVCVARC 600  
 QY 601 PSGVXPDLASYMPIMKFPDEEGACQCPINCTHSCVDLDKGGAPAEORASPLSIFNNFTV 660  
 DB 601 PSGVXPDLASYMPIMKFPDEEGACQCPINCTHSCVDLDKGGAPAEORASPLSIFNNFTV 660  
 QY 661 SFMLRVKVSASHLEKRRQOKIRKTYMRLLGETELVEBLTPSGAMPNQAQWRILKETEL 720  
 DB 661 SFMLRVKVSASHLEKRRQOKIRKTYMRLLGETELVEBLTPSGAMPNQAQWRILKETEL 720  
 QY 721 RRYKVLGSGAFGTYYKGIWI PDGENVKI PVAIKVLRRENTSPKANKELIDEAYMAGVGS 780  
 DB 721 RRYKVLGSGAFGTYYKGIWI PDGENVKI PVAIKVLRRENTSPKANKELIDEAYMAGVGS 780  
 QY 781 YVSRLLGICLTSTVQVLTQMLPVGCLLDHVRNENRGLSGODLNNCMQIAKMSYLEDR 840  
 DB 781 YVSRLLGICLTSTVQVLTQMLPVGCLLDHVRNENRGLSGODLNNCMQIAKMSYLEDR 840  
 QY 841 LVHRDLAARNVLYVSPNHVKITDPGLARLLIDETETHAAGKVPIMKMALESILRRFT 900  
 DB 841 LVHRDLAARNVLYVSPNHVKITDPGLARLLIDETETHAAGKVPIMKMALESILRRFT 900



Db 841 LVHRDLAARVIVKSPHVKITDFGLARLLDIDETEHNAQGVPIKMALESILRRRT 900  
 Qy 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPARISPDLEKXERLPPPICTIDVYIMVYKWM 960  
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKXERLPPPICTIDVYIMVYKWM 960  
 Qy 961 IDSCRRPFELVSEFSRMRADPQRFVIONEDLGPASPLDSTFFRSLEDDDMGLVDA 1020  
 Db 961 IDSCRRPFELVSEFSRMRADPQRFVIONEDLGPASPLDSTFFRSLEDDDMGLVDA 1020  
 Qy 1021 EBYLVPOQGFCDPAPAGAGMYHRRSSSTRSGGDLTLGLEPSEEAAPRSPAPSE 1080  
 Db 1021 EBYLVPOQGFCDPAPAGAGMYHRRSSSTRSGGDLTLGLEPSEEAAPRSPAPSE 1080  
 Qy 1081 AGSDVPFDGLGMAKAKLQSLPTHDESPLORYSEDPVPLPSETDGVVALTLTSPQPEY 1140  
 Db 1081 AGSDVPFDGLGMAKAKLQSLPTHDESPLORYSEDPVPLPSETDGVVALTLTSPQPEY 1140  
 Qy 1141 NOPDVRPQPPSPREGPLPARPAGATLERAKTLPQKNGVYKDVAFGAVENPEYLTQ 1200  
 Db 1141 NOPDVRPQPPSPREGPLPARPAGATLERAKTLPQKNGVYKDVAFGAVENPEYLTQ 1200  
 Qy 1201 GGAAPQHPPPAPSPAFDNLVWDQDPPERGAPSTFKGPTAENPEYGLDVPV 1255  
 Db 1201 GGAAPQHPPPAPSPAFDNLVWDQDPPERGAPSTFKGPTAENPEYGLDVPV 1255  
 RESULT 4  
 AAU74545  
 ID AAU74545 standard; Protein; 1255 AA.  
 AC AAU74545;  
 XX  
 DT 23-APR-2002 (first entry)  
 XX  
 DE Human HER2 (ErbB2) polypeptide.  
 XX  
 XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;  
 XX anti-ErbB antibody-maytansinoid conjugate; cancer; tumor; breast; ovary;  
 XX scromach; endometrium; salivary gland; lung; kidney; colon; colorectum;  
 XX thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;  
 XX glial disorder; astrocytal disorder; hypothalamic disorder;  
 XX glial disorder; macrophagal disorder; epithelial disorder;  
 XX stromal disorder; blastococic disorder; inflammatory disorder;  
 XX angiogenic disorder; immunological disorder.  
 XX  
 OS Homo sapiens.  
 XX  
 PN US2002001587-A1.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 16-MAR-2001; 2001US-0811123.  
 XX  
 PR 16-MAR-2000; 2000US-189844P.  
 PR 05-OCT-2000; 2000US-238327P.  
 XX  
 PA (ERIC/) ERICKSON S.  
 PA (SCHW/) SCHWALL R.  
 PA (SLIM/) SLIKOWSKI M.  
 XX  
 PI Erickson S, Schwall R, Slikowski M;  
 XX  
 DR WPI, 2002-163686/21.  
 DR N-PSDB; ABK14058.  
 XX  
 PT Treating tumour characterised by overexpression of epidermal growth  
 PT factor receptor, ErbB or cancer in mammal, comprises administering  
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal  
 XX  
 PS Example 3; Fig 7; 93pp; English.  
 CC The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor  
 CC receptor (ErbB) and does not respond or responds poorly, to treatment  
 CC with an anti-ErbB antibody, comprising administering to the mammal an  
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for  
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,  
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,  
 CC prostate and bladder, preferably breast cancer. The breast cancer is a  
 CC metastatic breast cancer or an aggressive form of metastatic breast  
 CC cancer which overexpresses ErbB2. The method is also useful for treating  
 CC neuronal, glial, astrocytal, hypothalamic, glandular, macrophagal,  
 CC epithelial, stromal, blastococic, inflammatory, angiogenic and  
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)  
 CC polypeptide of the invention.  
 XX  
 SQ Sequence 1255 AA;

Query Match 97.0%; Score 6627; DB 23; Length 1255;  
 Best Local Similarity 97.4%; Pred. No. 0;  
 Matches 1222; Conservative 6; Mismatches 27; Indels 0; Gaps 0;  
 Qy 1 MELALCWMGLLALLPQAASTQVCTGDMKRLPASPEHLDRLYQGCQVYQNL 60  
 Db 1 MELALCWMGLLALLPQAASTQVCTGDMKRLPASPEHLDRLYQGCQVYQNL 60  
 Qy 61 ELTYLPTNASLFLQDIOEVQGYLIANOVROYLQRLRYRGTLFEDNYALAVLNG 120  
 Db 61 ELTYLPTNASLFLQDIOEVQGYLIANOVROYLQRLRYRGTLFEDNYALAVLNG 120  
 Qy 61 ELTYLPTNASLFLQDIOEVQGYLIANOVROYLQRLRYRGTLFEDNYALAVLNG 120  
 Db 61 ELTYLPTNASLFLQDIOEVQGYLIANOVROYLQRLRYRGTLFEDNYALAVLNG 120  
 Qy 121 DELNNTPTVAGSPGGLRELOLRSLTEILKGVLIQRNPOLCYODTILMKDIFHKNOLA 180  
 Db 121 DELNNTPTVAGSPGGLRELOLRSLTEILKGVLIQRNPOLCYODTILMKDIFHKNOLA 180  
 Qy 181 LTLIDTNSRACHPSCSPKCKSGKSGESSBDCQSLTRVYCGGACRCGPLETDCHEQC 240  
 Db 181 LTLIDTNSRACHPSCSPKCKSGKSGESSBDCQSLTRVYCGGACRCGPLETDCHEQC 240  
 Qy 241 AAGCTGPRGSDCLALCFHFNHSGICELHPALVTVYTDFFESMPNEGRTVGASCVTACP 300  
 Db 241 AAGCTGPRGSDCLALCFHFNHSGICELHPALVTVYTDFFESMPNEGRTVGASCVTACP 300  
 Qy 301 YNYLSTDVGSCTLVCPFLHNOEYTAEDGTQRCCKSKPCARVYCYGIMQYIRANSKFIGIT 360  
 Db 301 YNYLSTDVGSCTLVCPFLHNOEYTAEDGTQRCCKSKPCARVYCYGIMQYIRANSKFIGIT 360  
 Qy 361 ELFEAGCKKIGSLAFLESPFDPASNTAALQPOLOVFTLEITGYLISAMPDSL 420  
 Db 361 ELFEAGCKKIGSLAFLESPFDPASNTAALQPOLOVFTLEITGYLISAMPDSL 420  
 Qy 421 DLVAFQNLQVIRGRILHNGAYSLTLOGIGISMLGLRSRLREGSGALIHNTTHLCFVHTV 480  
 Db 421 DLVAFQNLQVIRGRILHNGAYSLTLOGIGISMLGLRSRLREGSGALIHNTTHLCFVHTV 480  
 Qy 481 FMDOLFRRPQALLHTANRPEDECVGEGSLACHOLCARGHCHGPGPTQCVNCSQPLRGEC 540  
 Db 481 FMDOLFRRPQALLHTANRPEDECVGEGSLACHOLCARGHCHGPGPTQCVNCSQPLRGEC 540  
 Qy 541 VEECGVLOGLPREYVNAHCHLPCPECCOPONGSVTCFPGPADOCVACHYADPPCVARC 600  
 Db 541 VEECGVLOGLPREYVNAHCHLPCPECCOPONGSVTCFPGPADOCVACHYADPPCVARC 600  
 Qy 601 PSQVAFDLSYMPIMKFPDEBAGACOPCPINTCHSCVDLDDKCPAEGRASPLTISVSAVVG 660  
 Db 601 PSQVAFDLSYMPIMKFPDEBAGACOPCPINTCHSCVDLDDKCPAEGRASPLTISVSAVVG 660  
 Qy 661 SFWLRYVPVYASHEKRGQCKIRKYTWRRLLQETELVEPLTPSGAMPQAOARLKEHEL 720  
 Db 661 SFWLRYVPVYASHEKRGQCKIRKYTWRRLLQETELVEPLTPSGAMPQAOARLKEHEL 720  
 Qy 721 RKYVYLSGAGFYVYKGIWIPDGENVXIKVAILRENTSEKANKELIDEAYVAVAGVSP 780  
 Db 721 RKYVYLSGAGFYVYKGIWIPDGENVXIKVAILRENTSEKANKELIDEAYVAVAGVSP 780  
 Qy 781 YVSRLLGICLTSTVQLVTQLMPYGLCLDHRERNGRLGSDLLNWCQIAKXGMYLEDVR 840

```

Db 781 YVSRLLGICLTSTVQLTQMLPFGCLDHRKNGRSGSODLNMCMQIAKMSYLEDR 840
QY 841 LVHBDLAARNLVKSPNHVKITDFGLARLLDIDETEVHADQGVPIKMALESTLRRT 900
Db 841 LVHBDLAARNLVKSPNHVKITDFGLARLLDIDETEVHADQGVPIKMALESTLRRT 900
QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEGERLPOPICTIDVYIMVYCM 960
Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEGERLPOPICTIDVYIMVYCM 960
QY 961 IDSECRPFRELVESEFRMADPQRFVVIQNEDELGPASPLDSTFYRSLLEDMDGLVDA 1020
Db 961 IDSECRPFRELVESEFRMADPQRFVVIQNEDELGPASPLDSTFYRSLLEDMDGLVDA 1020
QY 1021 EBYLVPOGFCPPRAGAGGMVHRHRSSTRSGGDLTLGLPSESEAPRSPLASSEG 1080
Db 1021 EBYLVPOGFCPPRAGAGGMVHRHRSSTRSGGDLTLGLPSESEAPRSPLASSEG 1080
QY 1081 AGSDVPDGLGMAKAGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVPDGLGMAKAGLQSLPTHDPSPLOQYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NOPVRPQPPSPRSGPLPAAPAPATLEBRKTLSPGXGVYKVYAFAGAVENPEYLTPO 1200
Db 1141 NOPVRPQPPSPRSGPLPAAPAPATLEBRKTLSPGXGVYKVYAFAGAVENPEYLTPO 1200
QY 1201 GGAPQHPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGIDVYV 1255
Db 1201 GGAPQHPHPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYLGIDVYV 1255

```

## RESULT 5

AAW01111 standard; Protein; 1255 AA.

```

ID AAW01111 standard; Protein; 1255 AA.
AC AAW01111;
DT 01-JAN-1997 (first entry)
DE HER-2/neu protein.
XX HER-2/neu protein.
XX HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
XX breast cancer; ovary cancer; colon cancer; lung cancer;
XX prostate cancer; immunisation; tumour; vaccine; vector.
XX Homo sapiens.
OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX Domain 676..1255
XX /label= Intracellular domain
XX /note= "claimed domain, useful for immunisation"
XX
XX W09630514-A1.
XX
XX 03-OCT-1996.
XX
XX 28-MAR-1996; 96WO-US01689.
XX
XX 31-MAR-1995; 95US-0414417.
XX
XX (UNIW) UNIV WASHINGTON.
XX
XX Cheever MA, Disis ML;
XX
XX WPI: 1996-455361/45.
XX
XX N-PSDB: AAT140739.
XX
XX DNA encoding HER-2-neu poly(peptide(s)) - used for prevention or
XX treatment of malignancies with which the HER-2/neu oncogene is
XX associated
XX
XX Claim 2; Page 56-61; 71pp; English.

```

```

XX Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT140739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transfected host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;

```

```

Query Match 96.9%; Score 6623; DB 17; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

```

```

QY 1 MEIAALCRWGLLALLPFGMASTQVCTGDMKLRLPASPEYTHLMLRLHYGCCQVQVQNL 60
Db 1 MEIAALCRWGLLALLPFGMASTQVCTGDMKLRLPASPEYTHLMLRLHYGCCQVQVQNL 60
QY 61 ELTYLPTNASLSPLODIOEVQGVYLAHQVQVPLQRLIRVGTQLFEDNVALAVLNG 120
Db 61 ELTYLPTNASLSPLODIOEVQGVYLAHQVQVPLQRLIRVGTQLFEDNVALAVLNG 120
QY 121 DPANNTPTVGASPGGLRELQSLTEILKGVLIQRNPOLCYODTILMKDI FHKNNQLA 180
Db 121 DPANNTPTVGASPGGLRELQSLTEILKGVLIQRNPOLCYODTILMKDI FHKNNQLA 180
QY 181 LTLIDPNERACHPGSPMKSGRCWGESEDDQSILTRYCAGGACARCKPLPTDCCHEQC 240
Db 181 LTLIDPNERACHPGSPMKSGRCWGESEDDQSILTRYCAGGACARCKPLPTDCCHEQC 240
QY 241 AAGCTGPKHSDCLACLFENHSGICELHCPALVTYNTDFESMNPNEGRTYFASCVTACP 300
Db 241 AAGCTGPKHSDCLACLFENHSGICELHCPALVTYNTDFESMNPNEGRTYFASCVTACP 300
QY 301 YNYLSTDVQSCITLVCPDLHQEYVTAEDGTORCEKSPCARVCYGLGMQYIKANSKFIGIT 360
Db 301 YNYLSTDVQSCITLVCPDLHQEYVTAEDGTORCEKSPCARVCYGLGMQYIKANSKFIGIT 360
QY 361 ELFEAGCKKIFGSLAFIPSPFDGDPASNTAPLOPOLQVFEETLEETGLYISAMPDLSLP 420
Db 361 ELFEAGCKKIFGSLAFIPSPFDGDPASNTAPLOPOLQVFEETLEETGLYISAMPDLSLP 420
QY 421 DLSVFQNLQVIRGRIHNGAVSLTLQGLISWLGRLSRLBELSGLALIRHNTLCEVHTV 480
Db 421 DLSVFQNLQVIRGRIHNGAVSLTLQGLISWLGRLSRLBELSGLALIRHNTLCEVHTV 480
QY 481 PMDQLFRNHQALLHTANPEDECVGEGLACHQLCARHGCMWGPPTQCNCSQFLRGQEC 540
Db 481 PMDQLFRNHQALLHTANPEDECVGEGLACHQLCARHGCMWGPPTQCNCSQFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHECQFONGSVTCFGEADQCAVACHKDPFCVVARC 600
Db 541 VEECRVLOGLPREYVNAHCLPCHECQFONGSVTCFGEADQCAVACHKDPFCVVARC 600
QY 601 PSGVAPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGPAPQORASPLTSTFNFTV 660
Db 601 PSGVAPDLSYMPIMKFPDEEGACQPCPINCTHSCVDLDDKGPAPQORASPLTSTFNFTV 660
QY 661 SFMLRVPKVASHLEKROQKIRKTYMRLLOETELVEPLTPSGAMPNOAQRIKETEEL 720
Db 661 SFMLRVPKVASHLEKROQKIRKTYMRLLOETELVEPLTPSGAMPNOAQRIKETEEL 720
QY 721 RRYKVLGSGAFGTIVKGIWIPDGENVKIPVAIKVIRENTSPRANKIIDEAYVMAGVSP 780
Db 721 RRYKVLGSGAFGTIVKGIWIPDGENVKIPVAIKVIRENTSPRANKIIDEAYVMAGVSP 780
QY 781 YVSRLLGICLTSTVQLTQMLPFGCLDHRKNGRSGSODLNMCMQIAKMSYLEDR 840
Db 781 YVSRLLGICLTSTVQLTQMLPFGCLDHRKNGRSGSODLNMCMQIAKMSYLEDR 840

```

QY 841 LVHRDLAARNVAVKSPNHVKITDIFGLAALLDIDETEVHADGKVPKIMMALESLIRRRFT 900  
 Db 841 LVHRDLAARNVAVKSPNHVKITDIFGLAALLDIDETEVHADGKVPKIMMALESLIRRRFT 900  
 QY 901 HQSDVWSYGVTVWELMTFGAKPYDGPARE:PDLLKGERLPQPPICITIDVYIMVCKM 960  
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGPARE:PDLLKGERLPQPPICITIDVYIMVCKM 960  
 QY 961 IDSECRPRFRELVSFSRMADPQRFVITQNEDELGPASPLDSTYRSLLEDDMDLVDA 1020  
 Db 961 IDSECRPRFRELVSFSRMADPQRFVITQNEDELGPASPLDSTYRSLLEDDMDLVDA 1020  
 QY 1021 EEVLVPOGFCPPAPAGAGMHHRRSSSTRSGGGDLTGLPSESEAPRSLAPSEG 1080  
 Db 1021 EEVLVPOGFCPPAPAGAGMHHRRSSSTRSGGGDLTGLPSESEAPRSLAPSEG 1080  
 QY 1081 AGSDVFDLDMGAAGKIQSLPTHDPSFLQRYSDPTVPLPSETDGYVAPLTCSQPEYV 1140  
 Db 1081 AGSDVFDLDMGAAGKIQSLPTHDPSFLQRYSDPTVPLPSETDGYVAPLTCSQPEYV 1140  
 QY 1141 NOPDVRPQPSRPGSLPAPAPAGATLERAKTSPGKGVVQVAFAGAVENPEYLTPO 1200  
 Db 1141 NOPDVRPQPSRPGSLPAPAPAGATLERAKTSPGKGVVQVAFAGAVENPEYLTPO 1200  
 QY 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGIDVAV 1255  
 Db 1201 GGAAPQHPHPPAFSPAFDNLVYWDQDPPERGAPPSTFKGPTAENPEYLGIDVAV 1255

RESULT 6  
 AAW92406  
 ID AAW92406 standard; Protein; 1255 AA.

AAW92406;

21-APR-1999 (first entry)

Human HER-2/neu oncogene protein.

HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
 malignancy; treatment; tumour.

Homo sapiens.

Key Location/Qualifiers

Region 676..1255  
 /note="region which elicits immune response"

US5869445-A.

09-FEB-1999.

01-APR-1996; 96US-0625101.

01-APR-1996; 96US-0625101.

17-MAR-1993; 93US-0033644.

12-AUG-1993; 93US-0106112.

31-MAR-1995; 95US-0414417.

(UNIW) UNIV WASHINGTON.

Cheever MA, Disis ML;

WPI, 1999-152835/13.

N-PSDB; AAX01912.

Use of HER-2/neu polypeptides - for eliciting an immune response to

an HER-2/neu associated malignancy, particularly for treating or

preventing tumours

Claim 3; Column 31-38; 26pp; English.

This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune  
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and  
 CC B cells to produce an immune response to the HER-2/neu protein. The  
 CC method can be used for immunisation against a malignancy in which the  
 CC HER-2/neu oncogene is associated and in the treatment of an existing  
 CC tumour, or to prevent tumour occurrence or recurrence.

Sequence 1255 AA;

Query Match 96.9%; Score 6623; DB 20; Length 1255;  
 Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELAALCRNGLLALLPPGAASTQVCTGDMKRLPLASPTHLDMLRHLYOGQVQGNL 60  
 Db 1 MELAALCRNGLLALLPPGAASTQVCTGDMKRLPLASPTHLDMLRHLYOGQVQGNL 60  
 QY 61 ELTYLPTNASLSTFLQDIOEVQGYVLAHQVROVPLQRLRYRGTOLEFDNVALAVLDNG 120  
 Db 61 ELTYLPTNASLSTFLQDIOEVQGYVLAHQVROVPLQRLRYRGTOLEFDNVALAVLDNG 120  
 QY 121 DPUNNTTPVTGASPGGLRELQRLSTELIKGGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180  
 Db 121 DPUNNTTPVTGASPGGLRELQRLSTELIKGGVLIQRNPOLCYQDTILMKDIFHKNNOLA 180  
 QY 181 LTIIDNRSRACHPSCPKSGRCWSESSEDQSLTRTYCAGGCACRGLPDDCCHEOC 240  
 Db 181 LTIIDNRSRACHPSCPKSGRCWSESSEDQSLTRTYCAGGCACRGLPDDCCHEOC 240  
 QY 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTPESMNPGRATPFASCTYAC 300  
 Db 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTPESMNPGRATPFASCTYAC 300  
 QY 301 YNYLSTDVSGCTLVCPHNOEVTABEDGTORCEKSKPCARVCYGLGMEHLREYAVTSAN 360  
 Db 301 YNYLSTDVSGCTLVCPHNOEVTABEDGTORCEKSKPCARVCYGLGMEHLREYAVTSAN 360  
 QY 361 ELEFACCKKIFGSLATLPSPGDDPASNAPLOPBOLOVFEETLEETGYLYTSAMDLSLP 420  
 Db 361 ELEFACCKKIFGSLATLPSPGDDPASNAPLOPBOLOVFEETLEETGYLYTSAMDLSLP 420  
 QY 421 DLSVFQNTQVIRGRILHNGAYSILTLQGLISWLGRLSRLSGGLALIHNNHTLCFVATV 480  
 Db 421 DLSVFQNTQVIRGRILHNGAYSILTLQGLISWLGRLSRLSGGLALIHNNHTLCFVATV 480  
 QY 481 PMDQLFRNPHQALLHTANRPDECVSEGLAQQLCARHGWKPGPTQVCNCSQFLRGDEC 540  
 Db 481 PMDQLFRNPHQALLHTANRPDECVSEGLAQQLCARHGWKPGPTQVCNCSQFLRGDEC 540  
 QY 541 VEECRVLOGLPREYVVARHCLPCHPECCQPNQSVTCFGEADQCVACAHYKPPFCVANC 600  
 Db 541 VEECRVLOGLPREYVVARHCLPCHPECCQPNQSVTCFGEADQCVACAHYKPPFCVANC 600  
 QY 601 PSGVXPDLISYMPKWFPEDEGACQPCPINCTHSVDLDKCPAGBGRASPLTISFNNFTV 660  
 Db 601 PSGVXPDLISYMPKWFPEDEGACQPCPINCTHSVDLDKCPAGBGRASPLTISFNNFTV 660  
 QY 661 SFMLRPKYASASHLEBRQCKTRKXTMRLLQETELVEBVTSGAMPNQAQRILKETEL 720  
 Db 661 SFMLRPKYASASHLEBRQCKTRKXTMRLLQETELVEBVTSGAMPNQAQRILKETEL 720  
 QY 721 RRYKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPRANKELIDEAVYMGVGSF 780  
 Db 721 RRYKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTPRANKELIDEAVYMGVGSF 780  
 QY 781 YVSRLLIGLSTVQLVQMLPYGCLLDHVRNRGLSGODLLNCKMOTAKMSYLEDR 840  
 Db 781 YVSRLLIGLSTVQLVQMLPYGCLLDHVRNRGLSGODLLNCKMOTAKMSYLEDR 840  
 QY 841 LVHRDLAARNVAVKSPNHVKITDIFGLAALLDIDETEVHADGKVPKIMMALESLIRRRFT 900  
 Db 841 LVHRDLAARNVAVKSPNHVKITDIFGLAALLDIDETEVHADGKVPKIMMALESLIRRRFT 900

QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960  
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960  
 QY 961 IDSECRPRFRELVSFSSMARDPQRFVVIQNEIDIGPASPLDSTYRSLLEDDMDGLVDA 1020  
 Db 961 IDSECRPRFRELVSFSSMARDPQRFVVIQNEIDIGPASPLDSTYRSLLEDDMDGLVDA 1020  
 QY 1021 EBYLVPOQGFCCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEBEARSPLASSEG 1080  
 Db 1021 EBYLVPOQGFCCPDPAAGAGMWHRRSSSTRSGGDLTLGLEPSEBEARSPLASSEG 1080  
 QY 1081 AGSDVFDLDMGAAGKGLQSLPTHDPSPLOQYSEDPYPLPSETDGYAPLTCSPQPEYV 1140  
 Db 1081 AGSDVFDLDMGAAGKGLQSLPTHDPSPLOQYSEDPYPLPSETDGYAPLTCSPQPEYV 1140  
 QY 1141 NOPDVRFQPSPREGGLPAARPAAGATLERAKTSLPGKNGVYKDYAFAGAAVENEYTLTPQ 1200  
 Db 1141 NOPDVRFQPSPREGGLPAARPAAGATLERAKTSLPGKNGVYKDYAFAGAAVENEYTLTPQ 1200  
 QY 1201 GGAAFPQHPFPSPAFDNLVYWDQDPERGAAPSTFKGTPTAENPEYLGIDVPV 1255  
 Db 1201 GGAAFPQHPFPSPAFDNLVYWDQDPERGAAPSTFKGTPTAENPEYLGIDVPV 1255

## RESULT 7

AAB21198  
 ID AAB21198 standard; protein, 1255 AA.

XX AAB21198;

DT 12-JAN-2001 (first entry)

XX Human HER-2/neu proteoin.

KW Human; HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;

KW breast cancer; prostate cancer; ovarian cancer; lung cancer;

XX Homo sapiens.

XX WQ200044899-AL.

PD 03-AUG-2000.

XX 28-JAN-2000; 2000MO-US02164.

XX 29-JAN-1999; 99US-0117976.

PA (CORI-) CORIXA CORP.

PA (SMIK) SMITHKLINE BEECHAM.

PI Cheever MA, Gheysen D;

DR WPI: 2000-505976/45.

DR N-PSDB; AAA89736.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins

XX useful for vaccinating against breast, ovarian, colon, lung and

XX prostate cancers -

XX Claim 52; Fig 7; 128pp; English.

XX The present sequence is the human HER-2/neu protein. It is a member of  
 CC the tyrosine kinase family of receptor-like glycoproteins and shows  
 CC homology to the epidermal growth factor receptor (EGFR). It probably  
 CC plays a part in cell growth and/or differentiation. The HER-2/neu  
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a  
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
 CC domain may be used to treat or prevent cancer by eliciting or  
 CC enhancing an immune response to the HER-2/neu protein. It may be used  
 CC to treat malignancies such as breast, ovarian, colon, lung and  
 CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.  
 XX  
 SQ Sequence 1255 AA;

Query Match 96.9%; Score 6623; DB 21; Length 1255;  
 Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 METALACRNGLLALALPPGAASQVCTGDMKRLPASPETHLDMLRHLVYOGQVYQNL 60  
 Db 1 METALACRNGLLALALPPGAASQVCTGDMKRLPASPETHLDMLRHLVYOGQVYQNL 60  
 QY 61 ELTYLPTNASLSFLQDICEVQGVYLAHQVQVPLQRIIRIVRGTLFEDNYALAVLDNG 120  
 Db 61 ELTYLPTNASLSFLQDICEVQGVYLAHQVQVPLQRIIRIVRGTLFEDNYALAVLDNG 120  
 QY 121 DPLNNTPTVAGSPGGLRELQRLSLTEILKGVTLQRNQQLCYODTILMKDIFHKNNOLA 180  
 Db 121 DPLNNTPTVAGSPGGLRELQRLSLTEILKGVTLQRNQQLCYODTILMKDIFHKNNOLA 180  
 QY 181 LTLIDNRSRACHPCSPMKSGSRCWGESSEDCOSLTRVACAGGACRCKGPLPTDCHEQC 240  
 Db 181 LTLIDNRSRACHPCSPMKSGSRCWGESSEDCOSLTRVACAGGACRCKGPLPTDCHEQC 240  
 QY 241 AAGCTPRKASDCLACIHFHNSGICELHCPALVYNTDTRESMPNPEGRYTFGASCTYACP 300  
 Db 241 AAGCTPRKASDCLACIHFHNSGICELHCPALVYNTDTRESMPNPEGRYTFGASCTYACP 300  
 QY 301 YNLTSTDVSGCTLVCPLEHQEYTAEDGTQRCCKSPKARVCYGLGMEHLREVRATVSN 360  
 Db 301 YNLTSTDVSGCTLVCPLEHQEYTAEDGTQRCCKSPKARVCYGLGMEHLREVRATVSN 360  
 QY 361 ELEFACCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEETIGLYISAMPDSL 420  
 Db 361 ELEFACCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFETLEETIGLYISAMPDSL 420  
 QY 421 DLSPFNQLVITGRILHNGAVSLTQIGISWGLRSRLRELSGALILHNNHLCPVHTV 480  
 Db 421 DLSPFNQLVITGRILHNGAVSLTQIGISWGLRSRLRELSGALILHNNHLCPVHTV 480  
 QY 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCMGPGPTOCVNSQFLRGQEC 540  
 Db 481 PMDQLFRNPHQALLHTANRPEDECVGEGLAGHQLCARGHCMGPGPTOCVNSQFLRGQEC 540  
 QY 541 VEECRVLQCLPREYNNARCLPCHPECCQONGSYTCFGEALQCCVACAHYKDPFGVAC 600  
 Db 541 VEECRVLQCLPREYNNARCLPCHPECCQONGSYTCFGEALQCCVACAHYKDPFGVAC 600  
 QY 601 PSQVPRDLSTYMPIMKFPDEGACQCPICHTSCVYDLDDKGPAPORASPLTISAVVG 660  
 Db 601 PSQVPRDLSTYMPIMKFPDEGACQCPICHTSCVYDLDDKGPAPORASPLTISAVVG 660  
 QY 661 SFMLRVPKVSASHLEKRRQOKIRKTYMRRLLOETELVEPLTPSGAMPNOQRILKETEL 720  
 Db 661 SFMLRVPKVSASHLEKRRQOKIRKTYMRRLLOETELVEPLTPSGAMPNOQRILKETEL 720  
 QY 721 RYKYLGSAGFTVYKGIWIPDGEVVKIPVALKYLRNTPSKANELLDEAVVMGVSP 780  
 Db 721 RYKYLGSAGFTVYKGIWIPDGEVVKIPVALKYLRNTPSKANELLDEAVVMGVSP 780  
 QY 781 YVSRLLGICLTSTVQLVQLMPYGLLDHVENRRLSGDILNNCMQIAKMSVLEDDR 840  
 Db 781 YVSRLLGICLTSTVQLVQLMPYGLLDHVENRRLSGDILNNCMQIAKMSVLEDDR 840  
 QY 841 LVHRDLAANNVKSPNNHKTIDFGLARLLDDEHYADAGGVKIKMALESILRRFT 900  
 Db 841 LVHRDLAANNVKSPNNHKTIDFGLARLLDDEHYADAGGVKIKMALESILRRFT 900  
 QY 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960  
 Db 901 HQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICITIDVYIMVKKMM 960  
 QY 961 IDSECRPRFRELVSFSSMARDPQRFVVIQNEIDIGPASPLDSTYRSLLEDDMDGLVDA 1020

Db 961 IDSECRPRFRELVEFSPRMARDPQRFVVIQNEEDGPRASPLDSTFYSRLLEDGDLVDA 1020  
 QY 1021 EBYLVPOGGFFCPDPAPAGAGMVAHRRSSSTRSGGDLTLGLEPSEEAARSLAPSEG 1080  
 Db 1021 EBYLVPOGGFFCPDPAPAGAGMVAHRRSSSTRSGGDLTLGLEPSEEAARSLAPSEG 1080  
 QY 1081 AGSDVFPQGDLMGAAKGLQSLPTHDPSPLOQYSDDPVPLPSETDGVVAALTCSPQPEYV 1140  
 Db 1081 AGSDVFPQGDLMGAAKGLQSLPTHDPSPLOQYSDDPVPLPSETDGVVAALTCSPQPEYV 1140  
 QY 1141 NQPDVPRQPPSPREBPLPAARPAATLERAKTILSPGKGVKDVFAFGAVENPEYLTQ 1200  
 Db 1141 NQPDVPRQPPSPREBPLPAARPAATLERAKTILSPGKGVKDVFAFGAVENPEYLTQ 1200  
 QY 1201 GGAPQPHPPAPSPADNLYYNDQDPPEGAPSTFKGPTLAEPEYLGDLVAV 1255  
 Db 1201 GGAPQPHPPAPSPADNLYYNDQDPPEGAPSTFKGPTLAEPEYLGDLVAV 1255

## RESULT 8

ID AAY84780 standard; Protein; 1255 AA.  
 AC AAY84780;

DT 08-AUG-2000 (first entry)  
 XX

DE Amino acid sequence of the SPLICE erbB-2 receptor protein.

KW SPLICE erbB-2 receptor protein; cell transformation disorder; cancer;  
 tumor cell proliferation; tissue degeneration; atrophy;  
 bone resorption; inflammatory disease; degenerative disorder;  
 wound healing.

OS Homo sapiens.

PN WC0200020579-A1.

PD 13-APR-2000.

PF 01-OCT-1999; 99WO-CA00912.

PR 02-OCT-1998; 98US-0165192.

PA (UYMC-) UNIV MCMMASTER.

PI Muller WJ, Siegel PM;

DR WPI, 2000-303768/26.

DR N-PSDB; AAA14812.

PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE

PS erbB-2, inhibitors of the protein are useful for treatment of cancer -

PS Claim 3; Fig 2; 60pp; English.

CC The present sequence represents a SPLICE erbB-2 receptor protein. The  
 CC protein has an in-frame deletion of 16 amino acids, 2 of which are  
 CC conserved cysteine residues, compared to the unspliced protein. The  
 CC erbB-2 polynucleotide is used to construct probes for detecting  
 CC disorders of cell transformation such as cancer. Antibodies to the  
 CC protein may be used to detect SPLICE erbB-2 in a sample. Agents  
 CC (e.g., antisense oligonucleotides) which inhibit the expression of  
 CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and  
 CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful  
 CC for treating conditions involving damaged cells including conditions  
 CC in which degeneration of tissue occurs, such as atrophy, bone  
 CC resorption, inflammatory diseases, degenerative disorders of the  
 CC central nervous system and wound healing.

SQ Sequence 1255 AA;

Query Match 96.9%; Score 6623; DB 21; Length 1255;  
 Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELAAACRMGILLALIPGAASTVOCTGDMRLRLPASFTLDMRLHLYOGQVVOGML 60  
 Db 1 MELAAACRMGILLALIPGAASTVOCTGDMRLRLPASFTLDMRLHLYOGQVVOGML 60  
 QY 61 ELTYLPTNASLSFLQDIDQVGVYLAHNVQVPLQRLRIYRGTLFEDNALVALVDNG 120  
 Db 61 ELTYLPTNASLSFLQDIDQVGVYLAHNVQVPLQRLRIYRGTLFEDNALVALVDNG 120  
 QY 121 DPLANTPTTGASPGGLRELQRLSLTEILKGVLLQRPOLCYOITLWKDIFHKNOQA 180  
 Db 121 DPLANTPTTGASPGGLRELQRLSLTEILKGVLLQRPOLCYOITLWKDIFHKNOQA 180  
 QY 181 LTLIDPNRBRACHPCSPMKSGRCWGESESDQSILRTVCAGGACRCKPLPTDCHEQC 240  
 Db 181 LTLIDPNRBRACHPCSPMKSGRCWGESESDQSILRTVCAGGACRCKPLPTDCHEQC 240  
 QY 241 AAGCTGPKASDCLACLFENHSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300  
 Db 241 AAGCTGPKASDCLACLFENHSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300  
 QY 301 VNYLSTDVSSCTLVCPRLHNOEYTAADGTQRCCKSPCARVCYIGMOYIKANSKFGIT 360  
 Db 301 VNYLSTDVSSCTLVCPRLHNOEYTAADGTQRCCKSPCARVCYIGMOYIKANSKFGIT 360  
 QY 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELOVFEETLEITGLYLSAMPDSL 420  
 Db 361 ELEFAGCKKIFGSLAFLPESFDGDPASNTAPLQPELOVFEETLEITGLYLSAMPDSL 420  
 QY 421 DLIVQNTQVITRGLIANGAVSLTQSGISWLGIRSLRELGSLALIHNTLFCVHTV 480  
 Db 421 DLIVQNTQVITRGLIANGAVSLTQSGISWLGIRSLRELGSLALIHNTLFCVHTV 480  
 QY 481 FWDQLFRRPHOALHTANRPEDECVGEGLACHOC-CARGHCWPGPTQCVNCSQFLRGEC 540  
 Db 481 FWDQLFRRPHOALHTANRPEDECVGEGLACHOC-CARGHCWPGPTQCVNCSQFLRGEC 540  
 QY 541 VEECRVLTQGLPREVYNARHCLPCHBECCPONGSVTCFPEADQCAVACHYKDPFCVARC 600  
 Db 541 VEECRVLTQGLPREVYNARHCLPCHBECCPONGSVTCFPEADQCAVACHYKDPFCVARC 600  
 QY 601 PSQVPEPDLSTYPIKMFPEBEGACQPCPINCTHSCVDLDDKCPAEQRASPLTJISAVVG 660  
 Db 601 PSQVPEPDLSTYPIKMFPEBEGACQPCPINCTHSCVDLDDKCPAEQRASPLTJISAVVG 660  
 QY 661 SFWLKVPKVSASHLEKREOQKIRKYTMRLLOETELVEPLTPSGAMPVQAOMRLIKETEL 720  
 Db 661 SFWLKVPKVSASHLEKREOQKIRKYTMRLLOETELVEPLTPSGAMPVQAOMRLIKETEL 720  
 QY 721 RRVKYLGSAGATGYKGIWIPDGENVKIPVAIKYLRENTSPKANKEILDEAYVAVAGSP 780  
 Db 721 RRVKYLGSAGATGYKGIWIPDGENVKIPVAIKYLRENTSPKANKEILDEAYVAVAGSP 780  
 QY 781 YVSRLLGICLTSTQVLTQMLPQYGLDHPENGRGLSGQLLWMCQIAGMSYLEVR 840  
 Db 781 YVSRLLGICLTSTQVLTQMLPQYGLDHPENGRGLSGQLLWMCQIAGMSYLEVR 840  
 QY 841 LVHRLDLAARNLVSPNHVKITDFGLARLDDIDETVHAADGKVPIMMALESLIRRF 900  
 Db 841 LVHRLDLAARNLVSPNHVKITDFGLARLDDIDETVHAADGKVPIMMALESLIRRF 900  
 QY 901 HQSDVMSYGVVWELMTFGAKPYDGIPIAREIPDLLEGERLPCPITCTIDVYIMVCM 960  
 Db 901 HQSDVMSYGVVWELMTFGAKPYDGIPIAREIPDLLEGERLPCPITCTIDVYIMVCM 960  
 QY 961 IDSECRPRFRELVEFSPRMARDPQRFVVIQNEEDGPRASPLDSTFYSRLLEDGDLVDA 1020  
 Db 961 IDSECRPRFRELVEFSPRMARDPQRFVVIQNEEDGPRASPLDSTFYSRLLEDGDLVDA 1020  
 QY 1021 EBYLVPOGGFFCPDPAPAGAGMVAHRRSSSTRSGGDLTLGLEPSEEAARSLAPSEG 1080

```

DB 1021 EEYIVPQGFCCPBPAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1060
QY 1081 AGSDVFPDGLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGVAPLTCSPQPEYV 1140
DB 1081 AGSDVFPDGLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPPSPREGPLPAARPAAGATLEBAKTSPGKNGVYKVFARFGAVENPEYLTPO 1200
DB 1141 NOPDVRPQPPSPREGPLPAARPAAGATLEBAKTSPGKNGVYKVFARFGAVENPEYLTPO 1200
QY 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPERGAPPTKGTPTAENPEYLGADV 1255
DB 1201 GGAAPQHPPPAFSPAFDNLVYWDQDPERGAPPTKGTPTAENPEYLGADV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein, 1255 AA.
AC AAB85458;
DT 25-SEP-2001 (first entry)
DE Human HER-2/neu protein.
KM Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KM oncogene; cancer; cytostatic; vaccine; p185; C-erbB2.
XX Homo sapiens.
XX WO200153463-A2.
XX PD 26-JUL-2001.
XX PF 19-JAN-2001; 2001WO-US01850.
XX PR 21-JAN-2000; 2000US-0177545.
XX PA (CORI-) CORIXA CORP.
XX PI Cheever MA, Hand-Zimmermann S;
XX WI: 2001-476112/51.
XX N-PSDB; AAH23392.
XX DR N-PSDB; AAH23392.
XX PT New antigen-presenting cells, useful as vaccines for eliciting or
XX for enhancing an immune response to HER-2/neu protein, particularly useful
XX for treating or preventing cancer, e.g. breast cancer -
XX PS Claim 2; Page 41-46; 49pp; English.
XX CC The invention provides an isolated antigen-presenting cell, which
XX expresses at least an immunogenic portion of a polypeptide that produces
XX an immune response to HER-2/neu protein. The antigen-presenting cells are
XX useful as vaccines for eliciting or enhancing an immune response to
XX HER-2/neu protein, particularly in treating or preventing malignancies in
XX which the HER-2/neu oncogene is associated. Specifically, these are
XX useful for treating or preventing cancer, e.g. breast cancer, ovarian,
XX colon, lung or prostate cancers. The present sequence represents
XX the human HER-2/neu protein (also known as p185 or C-erbB2).
XX SO Sequence 1255 AA;

Query March 96.9%; Score 6623; DB 22; Length 1255;
Best local similarity 97.3%; Pred. No. 0;
Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELAALGRMGLLALLPFGAASVTCTGDKMLPASPEHLDMLHLYOGQVVGNTL 60
DB 1 MELAALGRMGLLALLPFGAASVTCTGDKMLPASPEHLDMLHLYOGQVVGNTL 60
QY 61 ELTYLPTNASLFLQDIDQEVGYVLIANQVRQVPLQRLRVGTQLFEDNYALAVLDNG 120

DB 61 ELTYLPTNASLFLQDIDQEVGYVLIANQVRQVPLQRLRVGTQLFEDNYALAVLDNG 120
QY 121 DPLANTTPTTGASPGGLBELQRLSITELKGVYLQRPOLCYOFTILMKOIFHNKOLA 180
DB 121 DPLANTTPTTGASPGGLBELQRLSITELKGVYLQRPOLCYOFTILMKOIFHNKOLA 180
QY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDDQSLTRFYVCAAGCARCKGAPLPTDCCHQC 240
DB 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDDQSLTRFYVCAAGCARCKGAPLPTDCCHQC 240
QY 241 AAGCTGPKRSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
DB 241 AAGCTGPKRSDCLACLFHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
QY 301 YNYLSTDPVSCCTLVCPHNOEYTAEDGTQRCCKSPCARVCYGLGMYIKANSKFIGIT 360
DB 301 YNYLSTDPVSCCTLVCPHNOEYTAEDGTQRCCKSPCARVCYGLGMYIKANSKFIGIT 360
QY 361 ELEFAGCKKIFGSLAFLESFDPDPASTAPLQPEQLQVFTLEBITGYLISAMPDSL 420
DB 361 ELEFAGCKKIFGSLAFLESFDPDPASTAPLQPEQLQVFTLEBITGYLISAMPDSL 420
QY 421 DLSVTQNTLVITGRILHNGAVSLTQIGISWLGRLSRLEUGSGLALHNTHLCPVHTV 480
DB 421 DLSVTQNTLVITGRILHNGAVSLTQIGISWLGRLSRLEUGSGLALHNTHLCPVHTV 480
QY 481 PMDOLFNRPHQALHTANRPEDECYEGELACHQLCARGHCMGPGFTQCVCNCSQFLRGEC 540
DB 481 PMDOLFNRPHQALHTANRPEDECYEGELACHQLCARGHCMGPGFTQCVCNCSQFLRGEC 540
QY 541 VEECRVYQLPREYVYNAHCLPCHEPCOPONGSVTCFGBEADQVACAHYKDPFCVARC 600
DB 541 VEECRVYQLPREYVYNAHCLPCHEPCOPONGSVTCFGBEADQVACAHYKDPFCVARC 600
QY 601 PSGVYKPDLSYMPIMKPEDEGACOPCPINCHSCVDLDDKGPAPORASPLTSINNFTV 660
DB 601 PSGVYKPDLSYMPIMKPEDEGACOPCPINCHSCVDLDDKGPAPORASPLTSINNFTV 660
QY 661 SFMLRVPKVSASHLEKROOKIRKTYMRLLOETELVBLTPSGAMPNOQMRILKETEL 720
DB 661 SFMLRVPKVSASHLEKROOKIRKTYMRLLOETELVBLTPSGAMPNOQMRILKETEL 720
QY 721 RKVKYLGSAGFGTYVKGITPDGENVKIPVAIKVRENTSPKAKVLEIDEAVYMGVSGP 780
DB 721 RKVKYLGSAGFGTYVKGITPDGENVKIPVAIKVRENTSPKAKVLEIDEAVYMGVSGP 780
QY 781 YYSRLIGICLTSTVOLVQMLPYGCLDHRNRRGLSGDLNKCQIAKMSYLEYVR 840
DB 781 YYSRLIGICLTSTVOLVQMLPYGCLDHRNRRGLSGDLNKCQIAKMSYLEYVR 840
QY 841 LVHRDLAARNVIVKSPNHVKITDFGLARLLDIDETRYHADGKVPKIMMALESILRRFT 900
DB 841 LVHRDLAARNVIVKSPNHVKITDFGLARLLDIDETRYHADGKVPKIMMALESILRRFT 900
QY 901 HOSDVSXVATVWELMTFEGAKPYDGPAREITDLEKGRILPOPICTIDVYMIWVKCM 960
DB 901 HOSDVSXVATVWELMTFEGAKPYDGPAREITDLEKGRILPOPICTIDVYMIWVKCM 960
QY 961 IDSECRPRELSESRMARDPORFVYIIONEDLPASPPLDSTFYRSLDEDDMDGLVDA 1020
DB 961 IDSECRPRELSESRMARDPORFVYIIONEDLPASPPLDSTFYRSLDEDDMDGLVDA 1020
QY 1021 EEYIVPQGFCCPBPAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
DB 1021 EEYIVPQGFCCPBPAGAGMWHHRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080
QY 1081 AGSDVFPDGLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGVAPLTCSPQPEYV 1140
DB 1081 AGSDVFPDGLGMAAKGLQSLPTHDPSPLOQYSEDPVPLPSETDGVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPPSPREGPLPAARPAAGATLEBAKTSPGKNGVYKVFARFGAVENPEYLTPO 1200
DB 1141 NOPDVRPQPPSPREGPLPAARPAAGATLEBAKTSPGKNGVYKVFARFGAVENPEYLTPO 1200

```

Db 1141 NQDVRPQPSREGPLPAARPAATLERPKTSLPGKGVKQVFAFGAVENPEYLTPO 1200  
 QY 1201 GGAAQPHPPAPSPADNLYYWDQDPERCAPSTKGTPTANPEYLLDVPV 1255  
 Db 1201 GGAAQPHPPAPSPADNLYYWDQDPERCAPSTKGTPTANPEYLLDVPV 1255

RESULT 10  
 AAG88267  
 ID AAG88267 standard; Protein; 1255 AA.  
 AC AAG88267;  
 DT 11-SEP-2001 (first entry)  
 DE HER2/neu amino acid sequence.  
 KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;  
 KM tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.  
 OS Homo sapiens.  
 PN MO200141787-A1.  
 PD 14-JUN-2001.  
 PF 11-DEC-2000; 2000WO-US33591.  
 PR 10-DEC-1999; 99US-0458299.  
 PA (EPIM-) EPIMUNE INC.  
 PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;  
 PI Keogh E;  
 XX WPI; 2001-374995/39.  
 DR An isolated prepared HER2/neu epitope useful in a vaccine for inducing  
 PT cellular immune responses for the prevention and treatment of cancer -  
 PS disclosure; Page 15; 199pp; English.

XX The present invention describes isolated prepared HER2/neu epitopes (I).  
 CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is  
 CC culture in vitro and binds to a complex of an epitope (I), bound to a  
 CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)  
 CC and a second epitope and the peptide is less than 50 contiguous amino  
 CC acids that have 100% identity with a native peptide sequence of HER2/neu;  
 CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical  
 CC excipient; (4) an isolated nucleic acid encoding (II); (I) has cytostatic  
 CC and immunostimulant activities, and can be used in vaccines. (I), (II)  
 CC and (III) are useful for inducing cellular immune responses for the  
 CC prevention and treatment of cancer. (I) and (II) are useful for  
 CC monitoring or evaluating an immune response to a tumour-associated  
 CC antigen when incubated with a T lymphocyte sample from a patient and  
 CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope  
 CC based vaccines mean that immunosuppressive epitopes that may be present  
 CC in whole antigens may be avoided. Selected epitopes may be combined to  
 CC enhance immunogenicity. The possible pathological side effects caused by  
 CC infectious agents or whole protein antigen is eliminated. The vaccine  
 CC provides the ability to direct and focus an immune response to multiple  
 CC selected antigens from the same pathogen. Epitope-based anti-tumour  
 CC vaccines provides the opportunity to combine epitopes derived from  
 CC multiple tumour-associated molecules addressing the problem of tumour-  
 CC tumour variability and reducing the likelihood of tumour escape due to  
 CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in  
 CC the exemplification of the present invention.

XX Sequence 1255 AA:  
 Query Match 96.9%; Score 6623; DB 22; Length 1255;

Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELAALCRWGLLALPFGAASQVCTGDMRLRLPASBETHLDMRLHYQCCQVQGNL 60  
 Db 1 MELAALCRWGLLALPFGAASQVCTGDMRLRLPASBETHLDMRLHYQCCQVQGNL 60

QY 61 ETTYLPTNASLSFLDIOEVQGYVLIANQVRQVPLQRIYRGTLQFEDNALAVLDNG 120  
 Db 61 ETTYLPTNASLSFLDIOEVQGYVLIANQVRQVPLQRIYRGTLQFEDNALAVLDNG 120

QY 121 DPLNNTPTVGTASPGGLRELQRLSTELIKGGVLIQBNPOLCYOTILMKDIFHKNOQA 180  
 Db 121 DPLNNTPTVGTASPGGLRELQRLSTELIKGGVLIQBNPOLCYOTILMKDIFHKNOQA 180

QY 181 LTLIDTNRBRACHPCSPMCKGSRGWSESDCSLTRVYACAGCARCKGPLPTDCHEQC 240  
 Db 181 LTLIDTNRBRACHPCSPMCKGSRGWSESDCSLTRVYACAGCARCKGPLPTDCHEQC 240

QY 241 AAGCTGPRKSDCLALCFPHNSGICELHCPALVTYNTDTPESMPNREGRTFGASCYTACP 300  
 Db 241 AAGCTGPRKSDCLALCFPHNSGICELHCPALVTYNTDTPESMPNREGRTFGASCYTACP 300

QY 301 YNYLSTDVSSCTLVCPPLHNOEYTAEDGTORCEKSCPCARVCYGLMOYIKANSKFIGIT 360  
 Db 301 YNYLSTDVSSCTLVCPPLHNOEYTAEDGTORCEKSCPCARVCYGLMOYIKANSKFIGIT 360

QY 361 ELTFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEBITGYLIASAPDSLIP 420  
 Db 361 ELTFAGCKKIFGSLAFLPESFDGDPASNTAPLQPEQLQVFTLEBITGYLIASAPDSLIP 420

QY 421 DLSVFQNLQVTRGRILHNGAVSLTLOGIGISWLGIRSLRELGSSALHNNHNLQFVHTV 480  
 Db 421 DLSVFQNLQVTRGRILHNGAVSLTLOGIGISWLGIRSLRELGSSALHNNHNLQFVHTV 480

QY 481 PMDQLFRPHQALLTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSQFDRGQEC 540  
 Db 481 PMDQLFRPHQALLTANRPEDECVGEGLACHQLCARGHCWGPPTQCVNCSQFDRGQEC 540

QY 541 VEECVLQGLPREYNAHCLPCHECQPNQSVTCFEPBADQCVACHYKDPPTCVARC 600  
 Db 541 VEECVLQGLPREYNAHCLPCHECQPNQSVTCFEPBADQCVACHYKDPPTCVARC 600

QY 601 PSQVPRDLSYPMKFPDEEGACQPCPINCCHSCVYDLDDKGPAPORASPLTIFNNFTV 660  
 Db 601 PSQVPRDLSYPMKFPDEEGACQPCPINCCHSCVYDLDDKGPAPORASPLTIFNNFTV 660

QY 661 SFMLRPVYASHEKRRQOKIRKYMRLQETELVBPPLPSGAMPNQAQWRIKETEL 720  
 Db 661 SFMLRPVYASHEKRRQOKIRKYMRLQETELVBPPLPSGAMPNQAQWRIKETEL 720

QY 721 RRVKVLGSGAGTYVKKGIWIPDGEVYKIPVAILKLRNTSKANKELIDEAYVWAGVSP 780  
 Db 721 RRVKVLGSGAGTYVKKGIWIPDGEVYKIPVAILKLRNTSKANKELIDEAYVWAGVSP 780

QY 841 LVHRDLARNTLVASPNHVKITTDGLARLDDIDETRYADGKVPKIMALESILRRFT 900  
 Db 841 LVHRDLARNTLVASPNHVKITTDGLARLDDIDETRYADGKVPKIMALESILRRFT 900

QY 901 HOSDWSAGVYVMEIMTFGAKYUGIPAREIPDLLEGGERLPPOPICTIDVYMMVWCMW 960  
 Db 901 HOSDWSAGVYVMEIMTFGAKYUGIPAREIPDLLEGGERLPPOPICTIDVYMMVWCMW 960

QY 961 IDSECRPRFRELVSSEFARMADPQRFVYIQNEDLGAPSLDSTFYRSLLEDDEDDGLVDA 1020  
 Db 961 IDSECRPRFRELVSSEFARMADPQRFVYIQNEDLGAPSLDSTFYRSLLEDDEDDGLVDA 1020

QY 1021 EBYLVPOGFCPCDPRAPAGAGMTHNRSSSTRGGDLTLGLEPSEEARPRSLASEG 1080







Db 1141 NQDVRPQPPSPRSGPLPAAPAGATLERPKTSLSPGKGVKDVFAFGAAVENEYLLTPQ 1200  
 Qy 1201 GGAAPQHPHPAFSPAFDNLVYWDQDPERGAPSPSTFKGTPTAENPEYLGLDVPV 1255  
 Db 1201 GGAAPQHPHPAFSPAFDNLVYWDQDPERGAPSPSTFKGTPTAENPEYLGLDVPV 1255

RESULT 12  
 AAE20479 standard; Protein; 1255 AA.  
 ID AAE20479;  
 AC AAE20479;  
 XX 01-JUL-2002 (first entry)  
 DT 01-JUL-2002 (first entry)  
 XX Human Her-2/neu protein.  
 DE Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;  
 KM human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Region 1021..1030  
 FT /note= "Naturally processed HLA-B44-restricted epitope"  
 XX WO200214503-A2.  
 XX 21-FEB-2002.  
 XX 14-AUG-2001; 2001WO-US41733.  
 XX 14-AUG-2000; 2000US-225153P.  
 PR 28-SEP-2000; 2000US-236428P.  
 PR 21-FEB-2001; 2001US-270520P.  
 XX (CORI-) CORIXA CORP.  
 PA Hand-Zimmermann S, Cheever MA, Foy TM, Lodes MT, Kalos MD;  
 PI McNeill PD, Vedvick TS;  
 DR WPI; 2002-280758/32.  
 DR N-PSDB; AAD32743.  
 XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,  
 PT prevention and diagnosis of cancer, preferably breast cancer  
 PS Disclosure; Page 114-117, 129pp; English.  
 XX The invention relates to an isolated Her-2/Neu polypeptide composition  
 CC effective for eliciting an immune response. The invention is useful for  
 CC eliciting an immune response in a patient where the patient is human  
 CC leukocyte antigen (HLA) B44 positive or is affected with breast cancer.  
 CC The composition is useful for the therapy and diagnosis of cancer.  
 CC Preferably breast cancer, in pharmaceutical compositions, e.g., vaccine  
 CC and other compositions for the diagnosis, prevention and treatment of  
 CC human malignancies, for stimulating and/or expanding T cells specific for  
 CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a  
 CC patient. The invention is useful for stimulating a T cell response in a  
 CC human patient, as probe or primer for nucleic acid hybridisation, to  
 CC selectively form duplex molecules with complementary stretches of the  
 CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full  
 CC length gene from a suitable library, and to direct expression of a  
 CC polypeptide in appropriate host cells. The composition is useful in  
 CC prophylactic or therapeutic applications and for the treatment of cancer,  
 CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-  
 CC associated malignancies. The invention is useful in gene therapy. The  
 CC present sequence is human Her-2/neu protein.  
 XX Sequence 1255 AA;  
 SQ Query Match 96.9%; Score 6623; DB 23; Length 1255;

Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;  
 Qy 1 MELAAACRMGGLLALLPFGAASQVCTGDMKRLPASPETHLDMRHLYOGQVQGNL 60  
 Db 1 MELAAACRMGGLLALLPFGAASQVCTGDMKRLPASPETHLDMRHLYOGQVQGNL 60  
 Qy 61 ELTYPTNASLSFLQDIOEVQGVYLAHNVQVPLQRLRYRGTLQFEDNVALAVLDNG 120  
 Db 61 ELTYPTNASLSFLQDIOEVQGVYLAHNVQVPLQRLRYRGTLQFEDNVALAVLDNG 120  
 Qy 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVYLIQNRNQLCYQDTILKQDFHKNQOLA 180  
 Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVYLIQNRNQLCYQDTILKQDFHKNQOLA 180  
 Qy 181 LTLIDNRRSACHPCSPMCKSGRCWSESESDCOSLTRVCAGGACARCKPPLPDCCHQC 240  
 Db 181 LTLIDNRRSACHPCSPMCKSGRCWSESESDCOSLTRVCAGGACARCKPPLPDCCHQC 240  
 Qy 241 AAGCTGPKASDCLACHFNHSGICEIHCALVTYNTDTFESMNPREGRYTFGASCVTACP 300  
 Db 241 AAGCTGPKASDCLACHFNHSGICEIHCALVTYNTDTFESMNPREGRYTFGASCVTACP 300  
 Qy 301 YNVLSTDVSGCTLVCPHNOEYTAEDGTQRCCKSPCARVCYGLGMOYIKANSKRTGIT 360  
 Db 301 YNVLSTDVSGCTLVCPHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVAVTSAN 360  
 Qy 361 ELSEFAGCKKIFGSLAFPEFSGDPAASNAPLQPEQLQVFETLEITGLYISAMPDSL 420  
 Db 361 IOEFAGCKKIFGSLAFPEFSGDPAASNAPLQPEQLQVFETLEITGLYISAMPDSL 420  
 Qy 421 DLSVFONLOVITGRILHNGAYSLTLQGLISWLGRLSRELSGLALIHNNHLCFVHTV 480  
 Db 421 DLSVFONLOVITGRILHNGAYSLTLQGLISWLGRLSRELSGLALIHNNHLCFVHTV 480  
 Qy 481 PMDQLFRNPHOALHTANRPEDECGEGLAGHQLCARHGWGPGTQCVNCSOFLRGQGC 540  
 Db 481 PMDQLFRNPHOALHTANRPEDECGEGLAGHQLCARHGWGPGTQCVNCSOFLRGQGC 540  
 Qy 541 VEECRVLOGLPREYVYNAHRLCPHCEQOPNGSVTCFGEADQCVAAHYKDPFCVANC 600  
 Db 541 VEECRVLOGLPREYVYNAHRLCPHCEQOPNGSVTCFGEADQCVAAHYKDPFCVANC 600  
 Qy 601 PEGVYKPDLSYMPIMKFPPEEGACQCPINCHSCVDLDKCGPAPQORASPLTISIYAVG 660  
 Db 601 PEGVYKPDLSYMPIMKFPPEEGACQCPINCHSCVDLDKCGPAPQORASPLTISIYAVG 660  
 Qy 661 SFMLRVPKVASHLEKRRQOKIRKXTMRLLQETELVEPLTPSGAMPQAOVRILKETEL 720  
 Db 661 ILLVVVLGVPIGLIKRRQOKIRKXTMRLLQETELVEPLTPSGAMPQAOVRILKETEL 720  
 Qy 721 RKVKYVLGSGAFSTVYKGIWIPDGENVKIPALIKVLRNTSPYANKELIDEAYVMAGVSP 780  
 Db 721 RKVKYVLGSGAFSTVYKGIWIPDGENVKIPALIKVLRNTSPYANKELIDEAYVMAGVSP 780  
 Qy 781 YVSRLLGICLTSTVOLQWMPYGLLHVHVENRRLSODLNMCMQIAKMSVLEDDR 840  
 Db 781 YVSRLLGICLTSTVOLQWMPYGLLHVHVENRRLSODLNMCMQIAKMSVLEDDR 840  
 Qy 841 LVHRDLAARNVLYKSPNVKITTDFGLARLLIDETEVYADGKVYIKMALESILRRFT 900  
 Db 841 LVHRDLAARNVLYKSPNVKITTDFGLARLLIDETEVYADGKVYIKMALESILRRFT 900  
 Qy 901 HOSDVMISGVYVWELMTGAPRYDIPAREIPDLLEKEERLPQPICTIDVYIMVYKCM 960  
 Db 901 HOSDVMISGVYVWELMTGAPRYDIPAREIPDLLEKEERLPQPICTIDVYIMVYKCM 960  
 Qy 961 IDSECRPRFREIVSEFSMARDPQRFVIVQNEIDGAPASPLSTFRLSLDDMDGDVDA 1020  
 Db 961 IDSECRPRFREIVSEFSMARDPQRFVIVQNEIDGAPASPLSTFRLSLDDMDGDVDA 1020  
 Qy 1021 EBYLVQGGFCPPDPAAGAMVHRHRSSSTRGGGDLTGLPSEBEARSPPLASEG 1080

Db 1021 EEVLVPOGPFPCDDPAPGAGMVRHRRSSSTRSGGDLTLGLEPSEEAAPRSLAPSEG 1080  
 QY 1081 AGSVFPQDGLGMAAKGIGSTPTHDPSLQRYSSDPVPLPSETDGYVALTCSPQPEYV 1140  
 Db 1081 AGSDVFDGDLGMAAKGIGSTPTHDPSLQRYSSDPVPLPSETDGYVALTCSPQPEYV 1140  
 QY 1141 NQPVPRQPPSPRSRGPPLPAARPAATLERAKTLSPKNGGVKDVFAFGAVENPEYLTPO 1200  
 Db 1141 NQPVPRQPPSPRSRGPPLPAARPAATLERAKTLSPKNGGVKDVFAFGAVENPEYLTPO 1200  
 QY 1201 GGAAPQHPHPPAFSPFNLVYWDPPERCAPSTRTGTATANPEYLGIDVY 1255  
 Db 1201 GGAAPQHPHPPAFSPFNLVYWDPPERCAPSTRTGTATANPEYLGIDVY 1255

## RESULT 13

AA051143  
 ID AA051143 standard; Protein; 1255 AA.  
 AC AA051143;  
 XX 17-JUN-2002 (first entry)  
 DT Human Her-2/neu oncogene-encoded p185 glycoprotein.  
 XX Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;  
 KW tyrosine kinase; receptor; c-erbB2; gene therapy.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Domain 1..653  
 FT Domain /note="extracellular domain"  
 FT Domain 676..1255  
 FT Domain /note="intracellular domain"  
 FT Domain 990..1255  
 FT /note="phosphorylation domain"  
 EN WO200212341-A2.  
 XX 14-FEB-2002.  
 PD 03-AUG-2001; 2001MO-US24283.  
 PF 03-AUG-2001; 2000US-0632507.  
 PR 03-AUG-2001; 2000US-0632507.  
 PA (CORI-) CORIXA CORP.  
 PA (SMK-) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX Cheever MA, Gheysen D;  
 PI WPI, 2002-241743/29.  
 DR N-PSDB; ABA92250.  
 DR Her-2/neu fusion protein for treating or preventing cancer by eliciting  
 PT or enhancing an immune response to the protein, has Her-2/neu  
 PT extracellular domain fused to Her-2/neu intracellular or  
 PT phosphorylation domain  
 XX Claim 68; Fig 7; 141pp; English.  
 PS The present sequence is that of human Her-2/neu (p185 glycoprotein  
 CC or c-erbB2), an oncogenic self-protein and target for anti-cancer  
 CC vaccines. The Her-2/neu gene is amplified and p185 is overexpressed  
 CC in a variety of cancers, including breast, ovarian, colon, lung and  
 CC prostate cancer. Her-2/neu is a member of the tyrosine kinase  
 CC family of receptor-like glycoproteins. It comprises an extracellular  
 CC domain with homology to the epidermal growth factor receptor  
 CC (EGFR), a highly hydrophobic transmembrane domain and a C-terminal  
 CC intracellular domain that also shows homology to EGFR. Its  
 CC overexpression correlates with a poor prognosis in breast and  
 CC ovarian cancers. The invention provides Her-2/neu fusion  
 CC proteins, nucleic acids encoding them, viral vectors, and vaccines

CC comprising the fusion proteins or nucleic acid molecules. In  
 CC preferred fusion proteins, the extracellular domain of a Her-2/neu  
 CC protein is fused to a Her-2/neu intracellular domain or  
 CC phosphorylation domain (or its Deltapart fragment). An immune  
 CC response to Her-2/neu protein is elicited or enhanced by  
 CC administering the fusion protein in the form of a vaccine, or by  
 CC transfecting cells of an animal *ex vivo* with a nucleic acid  
 CC encoding the fusion protein, and delivering the transfected cells  
 CC to the animal. The fusion proteins, nucleic acids, and isolated  
 CC specific T-cells are useful for inhibiting the development of a  
 CC cancer, especially breast, ovarian, colon, lung or prostate cancer  
 CC in a patient. T cells that specifically react with a Her-2/neu  
 CC fusion protein can be used to remove tumour cells from a sample in  
 CC order to inhibit the development of cancer in a patient.

## Sequence 1255 AA;

Query Match 96.9%; Score 6623; DB 23; Length 1255;  
 Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;

QY 1 MELAALCRMGILALALPPGAASQVCTGDMKRLRPADETHLDMRLHYOGGVQGNL 60  
 Db 1 MELAALCRMGILALALPPGAASQVCTGDMKRLRPADETHLDMRLHYOGGVQGNL 60  
 QY 61 ELTYLPTNASLSFLQDIQEVQGVYLIANQVAVPLQRLIRYRGTLQFEDNYALAVLNG 120  
 Db 61 ELTYLPTNASLSFLQDIQEVQGVYLIANQVAVPLQRLIRYRGTLQFEDNYALAVLNG 120  
 QY 121 DPLNNTPTVPGASPGGLRELQRLSTLEILKGGVLIQRPOLCYQDITLWMDIFPHKNOLA 180  
 Db 121 DPLNNTPTVPGASPGGLRELQRLSTLEILKGGVLIQRPOLCYQDITLWMDIFPHKNOLA 180  
 QY 181 LTLIDNRSRACHPCSPMKGSRCWSESSEDCSLTRTYCAGGCARCKBPLPTDCHEQC 240  
 Db 181 LTLIDNRSRACHPCSPMKGSRCWSESSEDCSLTRTYCAGGCARCKBPLPTDCHEQC 240  
 QY 241 AAGCTGPKMSDCLACHFNHSGICEALCALVTYNTDTFESMNPFGRTTFGASCVTACP 300  
 Db 241 AAGCTGPKMSDCLACHFNHSGICEALCALVTYNTDTFESMNPFGRTTFGASCVTACP 300  
 QY 301 YNLSLTPDVGCTLVCELNQVAVPLQRLIRYRGTLQFEDNYALAVLNG 360  
 Db 301 YNLSLTPDVGCTLVCELNQVAVPLQRLIRYRGTLQFEDNYALAVLNG 360  
 QY 361 ELFFAGCKKIFGSLAFLEPESFDGDPASNTAPLQPEQLQVETLEETIGLYISAWPDSL 420  
 Db 361 ELFFAGCKKIFGSLAFLEPESFDGDPASNTAPLQPEQLQVETLEETIGLYISAWPDSL 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISMGLSLBELSGGLLIHNTHLCEVHTV 480  
 Db 421 DLSVFQNLQVIRGRILHNGAYSLTLQGLGISMGLSLBELSGGLLIHNTHLCEVHTV 480  
 QY 481 PMDQFRNPHQALHTANRPEDECEVGEGLACHOLCARGHCMWGPPTQVCNCSQFLRGQEC 540  
 Db 481 PMDQFRNPHQALHTANRPEDECEVGEGLACHOLCARGHCMWGPPTQVCNCSQFLRGQEC 540  
 QY 541 VEECRVYQGLPREYVAVARHCLPCHPCQPNQSVTQFGPADQVCACAHKBPFCVAC 600  
 Db 541 VEECRVYQGLPREYVAVARHCLPCHPCQPNQSVTQFGPADQVCACAHKBPFCVAC 600  
 QY 601 PSQVKEPLSTMPIMKPPDEGACQPPINCTHSCVDLDKGCAPAEORASPLTIFNNFTV 660  
 Db 601 PSQVKEPLSTMPIMKPPDEGACQPPINCTHSCVDLDKGCAPAEORASPLTIFNNFTV 660  
 QY 661 SFULRVKVSASLHLEKRRQOKIRKYMRRLLQETELVEPLTPSGAMPNOQMRILKETEL 720  
 Db 661 SFULRVKVSASLHLEKRRQOKIRKYMRRLLQETELVEPLTPSGAMPNOQMRILKETEL 720  
 QY 721 RKRVKVGSGAFGVYGVYVPPDENKIPAIKVLBEKNSPKANKKILDEAYVAGVSGP 780  
 Db 721 RKRVKVGSGAFGVYGVYVPPDENKIPAIKVLBEKNSPKANKKILDEAYVAGVSGP 780

QY 781 YVSRLLGICLTSTVQLTQMLPFGCLLDHVRNRRGSLGSDLLNMCQIAKMSYLEDVR 840  
 Db 781 YVSRLLGICLTSTVQLTQMLPFGCLLDHVRNRRGSLGSDLLNMCQIAKMSYLEDVR 840  
 QY 841 LVHRDLAARVNLVKSPPNHVKITDFGLARLLDIDETEHADGGKVPIMKMALESILRRFT 900  
 Db 841 LVHRDLAARVNLVKSPPNHVKITDFGLARLLDIDETEHADGGKVPIMKMALESILRRFT 900  
 QY 901 HQSDVMSYGVTVWELMTFGAKPYPDGIIPAREIPDLLEKGERLPQPICTIDVYMIWKKCM 960  
 Db 901 HQSDVMSYGVTVWELMTFGAKPYPDGIIPAREIPDLLEKGERLPQPICTIDVYMIWKKCM 960  
 QY 961 IDSECRPRFRELSEFSRMAKDPORFVIVIONEDLGPASPLDSTFYSLIEDDDMDGLVDA 1020  
 Db 961 IDSECRPRFRELSEFSRMAKDPORFVIVIONEDLGPASPLDSTFYSLIEDDDMDGLVDA 1020  
 QY 1021 EBYLVPOQGFCCDPAPAGAGVYVHRRSSSTRSGGDDTLGLEPSEEAAPRSLAPSEG 1080  
 Db 1021 EBYLVPOQGFCCDPAPAGAGVYVHRRSSSTRSGGDDTLGLEPSEEAAPRSLAPSEG 1080  
 QY 1081 AGSDVDFDGLGMAAGKLOSLPTHDSPLQRYSEDPVLPBSETDGYVAPLTCSPPPEYV 1140  
 Db 1081 AGSDVDFDGLGMAAGKLOSLPTHDSPLQRYSEDPVLPBSETDGYVAPLTCSPPPEYV 1140  
 QY 1141 NQPDVAPQPPSPREGPLPARPAGATLEBAKTLSPGKNGVYVDFAFGAVENPEYLTPQ 1200  
 Db 1141 NQPDVAPQPPSPREGPLPARPAGATLEBAKTLSPGKNGVYVDFAFGAVENPEYLTPQ 1200  
 QY 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPTFFKGTPTANPEYVGLDVPV 1255  
 Db 1201 GGAAPQHPPPAPSPAFDNLVYWDQDPPERGAPPTFFKGTPTANPEYVGLDVPV 1255

RESULT 14  
 AAU77114  
 ID AAU77114 standard; Protein, 1255 AA.  
 XX  
 AC AAU77114;  
 XX  
 DT 05-JUN-2002 (first entry)  
 XX  
 DE Human Her-2/neu polypeptide.  
 XX  
 KW Human; Her-2/neu; cytostatic; haematological malignancy; CML;  
 KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;  
 KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;  
 KW Hodgkin's lymphoma; T cell therapy.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200213847-A2.  
 XX  
 PD 21-FEB-2002.  
 XX  
 PF 13-AUG-2001; 2001WO-US25408.  
 XX  
 PR 14-AUG-2000; 2000US-0638280.  
 PR 28-SEP-2000; 2000US-0675904.  
 XX  
 PA (CORI-) CORIXA CORP.  
 XX  
 PI Gaiger A, Cheever MA, Hand-Zimmermann S,  
 XX WPI; 2002-280741/32.  
 DR N-PSDB; ABLI0730.  
 XX  
 PT Inhibiting haematological malignancy development by administering  
 PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide  
 PT encoding the polypeptide, or antigen presenting cells expressing the  
 PT polypeptide  
 XX  
 PS Disclosure; Page 71-74; 74p; English.

CC The invention relates to a method for inhibiting development of  
 CC haematological malignancy in a patient by administering a polypeptide  
 CC comprising an immunogenic portion of Her-2/neu or a polynucleotide  
 CC encoding the polypeptide. Antigen presenting cells that express the  
 CC protein can also be administered. The sequences are used for inhibiting  
 CC development of haematological malignancy such as acute myelogenous  
 CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic  
 CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's  
 CC lymphoma. This sequence represents the human Her-2/neu polypeptide.  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 96.9%; Score 6623; DB 23; Length 1255;  
 Best Local Similarity 97.3%; Freq. No. 0;  
 Matches 1221; Conservative 6; Mismatches 28; Indels 0; Gaps 0;  
 QY 1 MELALCMGILLALPPGAASVCTGTDKMLRLPAPETHLMDLRLHYQCCVYQGNL 60  
 Db 1 MELALCMGILLALPPGAASVCTGTDKMLRLPAPETHLMDLRLHYQCCVYQGNL 60  
 QY 61 ELYVPTNASLSFLQDIEVQGYVLIANQVQVPLQRLIVRGTLQFEDNYVALVLDNG 120  
 Db 61 ELYVPTNASLSFLQDIEVQGYVLIANQVQVPLQRLIVRGTLQFEDNYVALVLDNG 120  
 QY 121 DPLNNTTGTASPGGLRELQRLSLTELKGVLIQRNPOLCYQDTILMKDIFHNKQGLA 180  
 Db 121 DPLNNTTGTASPGGLRELQRLSLTELKGVLIQRNPOLCYQDTILMKDIFHNKQGLA 180  
 QY 181 LTLIDTNRSAHPSPCKSRGSESSDCCSLTRVCAAGCARGKPLPTDCCHQC 240  
 Db 181 LTLIDTNRSAHPSPCKSRGSESSDCCSLTRVCAAGCARGKPLPTDCCHQC 240  
 QY 241 AAGCTGPRASDCLAFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300  
 Db 241 AAGCTGPRASDCLAFHNHSGICELHCPALVTYNTDTFESMPNPEGRYTFGASCVTAC 300  
 QY 301 YNYLSTDVSGCTVCPHNCVTAEDGQRECKSKPARVCYGLGMYIANSFRTIT 360  
 Db 301 YNYLSTDVSGCTVCPHNCVTAEDGQRECKSKPARVCYGLGMYIANSFRTIT 360  
 QY 361 EIEFAGCKKIFGSLAFPSDFDGPASNTAPLQPEOLQVFETLEITGYLISAMPDLP 420  
 Db 361 EIEFAGCKKIFGSLAFPSDFDGPASNTAPLQPEOLQVFETLEITGYLISAMPDLP 420  
 QY 421 DLSVFQNLQVIRGILHNGASLTLOIGIGISWLGRLSRLREGSGALIHNTHLCPVHTV 480  
 Db 421 DLSVFQNLQVIRGILHNGASLTLOIGIGISWLGRLSRLREGSGALIHNTHLCPVHTV 480  
 QY 481 PMDQFRNPHQALHTANRPEDECVEGLACHOLCARGHCMPGPTQCVNSQFLRGQEC 540  
 Db 481 PMDQFRNPHQALHTANRPEDECVEGLACHOLCARGHCMPGPTQCVNSQFLRGQEC 540  
 QY 541 VEECRVILQGLPREYVNAHCLPCHPECOQONGSVTCFPEADQCVACHYKDPFCVARC 600  
 Db 541 VEECRVILQGLPREYVNAHCLPCHPECOQONGSVTCFPEADQCVACHYKDPFCVARC 600  
 QY 601 PSQVAPDLSYMPKPFDEBGACQPCPINTCHSCVDLDDKCPAEQASPLTSTIISAVVG 660  
 Db 601 PSQVAPDLSYMPKPFDEBGACQPCPINTCHSCVDLDDKCPAEQASPLTSTIISAVVG 660  
 QY 661 SFWLRVPSVASHLEKRRQOKIRKYTRRLQETELVEPLTPSGAMPNOQAMRLIKETEL 720  
 Db 661 SFWLRVPSVASHLEKRRQOKIRKYTRRLQETELVEPLTPSGAMPNOQAMRLIKETEL 720  
 QY 721 RKVAVLSSGAFGTYYKGIWIPDGENVKIPVAIKLRENTSPKANKELIDBAVYAGVSP 780  
 Db 721 RKVAVLSSGAFGTYYKGIWIPDGENVKIPVAIKLRENTSPKANKELIDBAVYAGVSP 780  
 QY 781 YVSRLLGICLTSTVQLTQMLPFGCLLDHVRNRRGSLGSDLLNMCQIAKMSYLEDVR 840  
 Db 781 YVSRLLGICLTSTVQLTQMLPFGCLLDHVRNRRGSLGSDLLNMCQIAKMSYLEDVR 840  
 QY 841 LVHRDLAARVNLVKSPPNHVKITDFGLARLLDIDETEHADGGKVPIMKMALESILRRFT 900

Db 841 LVHRDLAARNVLVKS... 900  
 Qy 901 HQSDVWSYGVTVWELMTFGAKPYDGI... 960  
 Db 901 HQSDVWSYGVTVWELMTFGAKPYDGI... 960  
 Qy 961 IDSECRPRFRELVESEFSRMARDP... 1020  
 Db 961 IDSECRPRFRELVESEFSRMARDP... 1020  
 Qy 1021 EBYLVPOQGFPCPDPAAGAGM... 1080  
 Db 1021 EBYLVPOQGFPCPDPAAGAGM... 1080  
 Qy 1081 AGSDVFPDGLGMAKAGLQSL... 1140  
 Db 1081 AGSDVFPDGLGMAKAGLQSL... 1140  
 Qy 1141 NQPDVPRQPPSPREBGLPAAR... 1200  
 Db 1141 NQPDVPRQPPSPREBGLPAAR... 1200  
 Qy 1201 GGAAPOHPHPPAPSPAFDNL... 1255  
 Db 1201 GGAAPOHPHPPAPSPAFDNL... 1255

## RESULT 15

ID AAR39568 standard; Protein: 1433 AA.  
 AAR39568;

AC AAR39568;  
 DT 07-FEB-1994 (first entry)  
 DE Sequence of c-erbB-2 tumour antigen.  
 KM Tumour antigen; c-erbB-2; glycoprotein.  
 OS Homo sapiens.  
 PN MO3316185-A.  
 PD 19-AUG-1993.  
 PF 05-FEB-1993; 93WO-US01055.  
 PR 06-FEB-1992; 92US-0831967.  
 PA (CETU) CETUS ONCOLOGY CORP.  
 PA (CREA-) CREATIVE BIOMOLECULES INC.  
 PL Houston TX, Huston JS, Oppermann H, Ring DB,  
 DR WPI: 1993-272889/34.  
 DR N-PSDB; AAQ46083.  
 PT New single chain Fv polypeptide binding to C-erbB-2 tumour  
 PT antigen - for imaging or treating breast or ovarian cancer etc.  
 PS Disclosure; pages 48-54; 87pp; English.  
 CC c-erbB-2 refers to a protein antigen expressed on the surface of  
 CC tumour cells, such as breast and ovarian tumour cells, which is an  
 CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric  
 CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39583 represents  
 CC the location of a stop codon in AAQ46083.  
 SQ Sequence 1433 AA;

Query Match 96.3%; Score 6580; DB 14; Length 1433;  
 Best Local Similarity 96.7%; Pred. No. 0;

Matches 1214; Conservative 8; Mismatches 33; Indels 0; Gaps 0;

Qy	1	MELALCRKGLIALLPQAASTGYCTGTDMLRPAPEHLLDMLRHLYOGCCQVYQNL	60
Db	1	MELALCRKGLIALLPQAASTGYCTGTDMLRPAPEHLLDMLRHLYOGCCQVYQNL	60
Qy	61	ELTYLPTNASLSFLDIOEVQGYVLIANQVQVLPQRLRIYRGTLFEDNYALAVLNG	120
Db	61	ELTYLPTNASLSFLDIOEVQGYVLIANQVQVLPQRLRIYRGTLFEDNYALAVLNG	120
Qy	121	DELNNTPTVAGSPGRLRELQRLTELLKGVLIQRNPOLCYQDTILMKDIFHNQOLA	180
Db	121	DELNNTPTVAGSPGRLRELQRLTELLKGVLIQRNPOLCYQDTILMKDIFHNQOLA	180
Qy	181	LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVAGGACRCKPLPTDCHEQC	240
Db	181	LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVAGGACRCKPLPTDCHEQC	240
Qy	241	AACTGPKHSDCLAHFNHSGICELHCPALVTYNTDFESNPNBEGRTYTCASCVTACP	300
Db	241	AACTGPKHSDCLAHFNHSGICELHCPALVTYNTDFESNPNBEGRTYTCASCVTACP	300
Qy	301	YNYLSTDVSGCTLVCPHNOEYTAEDGTORCEKSPCARVCYGLMOYIKANSKFIGIT	360
Db	301	YNYLSTDVSGCTLVCPHNOEYTAEDGTORCEKSPCARVCYGLMOYIKANSKFIGIT	360
Qy	361	ELFPAQCKKI FGSIAFLPESEFDGDPASNTAPLQBPOLQVFTLEBITGYLISAMPDILP	420
Db	361	ELFPAQCKKI FGSIAFLPESEFDGDPASNTAPLQBPOLQVFTLEBITGYLISAMPDILP	420
Qy	421	DISVQNLQVIGRIHLNGAYSLTQGLISWLGRLSRLBELSGSLALIHNNHLSFVHTV	480
Db	421	DISVQNLQVIGRIHLNGAYSLTQGLISWLGRLSRLBELSGSLALIHNNHLSFVHTV	480
Qy	481	PMDQLFRNPHQALHTANRPEDECGEGIAHQCLCARGCMGPCTQCVCNSQFLRGQEC	540
Db	481	PMDQLFRNPHQALHTANRPEDECGEGIAHQCLCARGCMGPCTQCVCNSQFLRGQEC	540
Qy	541	VEECRYLQGLPREYVNAHCHLPCHPECCOPNGSVTCFGBEADQCAACHYKDPFCVARG	600
Db	541	VEECRYLQGLPREYVNAHCHLPCHPECCOPNGSVTCFGBEADQCAACHYKDPFCVARG	600
Qy	601	PSGVPRDLSYMPIMKFPDEEGACQPCPINCTHSCVLDLDDKCPAQGRASPSTISAVVG	660
Db	601	PSGVPRDLSYMPIMKFPDEEGACQPCPINCTHSCVLDLDDKCPAQGRASPSTISAVVG	660
Qy	661	SFMLRVKVSASHLEKRRQOKIRKYMRLLOETELVBLTPSGAMPQAOQRILKETEL	720
Db	661	SFMLRVKVSASHLEKRRQOKIRKYMRLLOETELVBLTPSGAMPQAOQRILKETEL	720
Qy	721	RKYKTLGSGAEGTYVKGIMIPDEGVNKIPVAIKYLRNTSPYANKEILDEAYVMGVGSP	780
Db	721	RKYKTLGSGAEGTYVKGIMIPDEGVNKIPVAIKYLRNTSPYANKEILDEAYVMGVGSP	780
Qy	781	YVSRLLGICLTSTVOLVQMLPYGCLLDHVENRGRLSGQDLNMCQIAKMSYLEDR	840
Db	781	YVSRLLGICLTSTVOLVQMLPYGCLLDHVENRGRLSGQDLNMCQIAKMSYLEDR	840
Qy	841	LVHRDLAARNVLVKS... 900	
Db	841	LVHRDLAARNVLVKS... 900	
Qy	901	HQSDVWSYGVTVWELMTFGAKPYDGI... 960	
Db	901	HQSDVWSYGVTVWELMTFGAKPYDGI... 960	
Qy	961	IDSECRPRFRELVESEFSRMARDP... 1020	
Db	961	IDSECRPRFRELVESEFSRMARDP... 1020	
Qy	1021	EEYLVPOQGFPCPDPAAGAGM... 1080	
Db	1021	EEYLVPOQGFPCPDPAAGAGM... 1080	

```
QY 1081 AGSDVFDGDLGMAKGIQSLPTHDPSPLOKRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
Db 1081 AGSDVFDGDLGMAKGIQSLPTHDPSPLOKRYSEDPVPLPSETDGYVAPLTCSPQPEYV 1140
QY 1141 NOPDVRPQPSREGPLPAPAPAGATLERAKTLPKXGKGVKDVFAFGAVENPEYLLTPQ 1200
Db 1141 NOPDVRPQPSREGPLPAPAPAGATLERAKTLPKXGKGVKDVFAFGAVENPEYLLTPQ 1200
QY 1201 GGAAPOPHPPPAFSPAFDNLVYWDODPPERGAPPSFTKGTPTAENPEYLGIDVPV 1255
Db 1201 GGAAPOPHPPPAFSPAFDNLVYWDODPPERGAPPSFTKGTPTAENPEYLGIDVPV 1255
```

Search completed: July 22, 2003, 09:16:28  
Job time : 41.9339 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.25 Seconds  
(without alignments)  
5413.772 Million cell updates/sec

Title: SEQ4-653-675-14

Perfect score: 6814  
Sequence: 1 MELAALCRMLLALP...TFKGTPTANPEYLGLDVPV 1253

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database:

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6604	96.9	1255	1 A24571	protein-tyrosine k
2	5811	85.3	1260	1 TVRTNU	protein-tyrosine k
3	5801.5	85.1	1254	2 I48161	p-185 precursor -
4	3659	45.0	1210	1 GQHU	epidermal growth f
5	3640	44.6	1210	2 A53183	epidermal growth f
6	3017.5	44.3	1223	1 TVGHV	epidermal growth f
7	2908.5	42.7	1308	2 A47253	epidermal growth f
8	2598	38.1	1166	1 S6142	protein-tyrosine k
9	2380.5	34.9	1342	2 A36223	kinase-related tra
10	2292.5	33.6	1339	2 JC4387	epidermal growth f
11	1688.5	24.8	698	1 TVFVU	protein-tyrosine k
12	1626	23.9	604	1 TVYUH	protein-tyrosine k
13	1599.5	23.5	1330	1 GQFE	epidermal growth f
14	1570	22.0	544	2 S3745	protein-tyrosine k
15	1563	22.9	545	2 S00727	kinase-related tra
16	1546	22.7	540	2 B44776	protein-tyrosine k
17	1544	22.7	540	1 TVFVB	protein-tyrosine k
18	1509	22.1	644	2 A36325	epidermal growth f
19	1487	18.9	1323	2 E88257	protein let-23 [im
20	1487	18.9	1374	2 S70712	protein-tyrosine k
21	1194	17.5	1369	2 S70713	protein-tyrosine k
22	1142	16.8	1717	1 A45558	epidermal growth f
23	1126	16.5	527	2 A42032	epidermal growth f
24	977.5	14.3	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	724	10.6	1363	2 T43220	insulin-like growt
28	706	10.4	1383	2 A36090	insulin receptor p
29	705.5	10.4	1372	2 A34157	insulin receptor p

30	705	10.3	1382	1 INHUR	insulin receptor p
31	691.5	10.1	1300	2 A36502	insulin receptor-r
32	685	10.1	1477	2 T18534	protein-tyrosine k
33	680	10.0	1607	2 T43212	insulin-like growt
34	667	9.8	1268	2 B36502	insulin receptor-r
35	638	9.4	1367	1 IGHUR1	insulin-like growt
36	631.5	9.3	2148	1 A56081	insulin receptor -
37	625	9.2	2101	2 S57245	insulin receptor (
38	624	9.2	1371	2 A33837	insulin-like growt
39	622	9.1	1390	2 T10346	insulin receptor -
40	604	8.9	1114	1 S05582	protein-tyrosine k
41	592	8.7	987	2 A54092	protein-tyrosine k
42	584.5	8.6	952	2 I50612	protein-tyrosine k
43	584.5	8.6	1091	2 S33596	protein-tyrosine k
44	576.5	8.5	984	2 A39753	protein-tyrosine k
45	572.5	8.4	801	4 TVHURE	transforming prote

## ALIGNMENTS

RESULT 1  
A24571  
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human  
N:Alternate names: C-erb-B-2 protein precursor; Kinase-related transforming proteir  
C:Species: Homo sapiens (man)  
C>Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999  
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622  
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Senda, K.; Nomura, N.; Miyajima, N.; Saito, N.  
Nature 319, 230-234, 1986  
A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal gr  
A:Reference number: A24571; MUID:86118653; PMID:3003577  
A:Accession: A24571  
A:Molecule type: mRNA  
A:Residues: 1-1255 <YAM>  
A:Cross-references: GB:X03163; NID:G31197; PIDN:CAA27060.1; PID:G31198  
R:Senda, K.; Kanata, N.; Toyoshima, K.; Yamamoto, T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6487-6501, 1985  
A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/ge  
A:Reference number: A25491; MUID:86016729; PMID:2995967  
A:Accession: A25491  
A:Molecule type: DNA  
A:Residues: 717-1031 <SEM>  
A:Cross-references: GB:M1167; NID:G182163; PIDN:AAA35808.1; PID:G553282  
R:Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeb  
Science 230, 1132-1139, 1985  
A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares c  
A:Reference number: A44188; MUID:86070181; PMID:2999974  
A:Accession: A44188  
A:Molecule type: DNA  
A:Residues: 740-910 <COU1>  
A:Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989  
A:Accession: B44188  
A:Molecule type: mRNA  
A:Residues: 1-517; RALU, 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>  
A:Cross-references: GB:M11730; NID:G183986  
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.  
Science 229, 974-976, 1985  
A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma  
A:Reference number: I59509; MUID:85272597; PMID:2992089  
A:Accession: I59509  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 332-309 <REX>  
A:Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808  
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Girol, D.  
Mol. Cell. Biol. 7, 2597-2601, 1987  
A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcripti  
A:Reference number: I57622; MUID:87286898; PMID:3039351  
A:Accession: I57622  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-191 <TAL>

F:558-680/Domain: transmembrane #status predicted <TM>  
 F:723-988/Domain: protein kinase homology <KIN>  
 F:731-739/Region: protein kinase ATP-binding motif  
 F:11,191,263,535,576,634/Binding site: carboxylate (Asn) (covalent) #status predicted  
 F:591/Binding site: phosphate (Thr) (covalent) #status predicted  
 F:758/Active site: Lys #status predicted  
 F:682,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.3%; Score 5811; DB 1; Length 1260;  
 Best Local Similarity 85.2%; Pred. No. 5, 1e-232;  
 Matches 1074; Conservative 57; Mismatches 118; Indels 12; Gaps 4;

QY 1 MELALCRWGLLALLPFGAASVCTGTDMKRLPASPEHLDMLRHLVYGCQVYQNL 60  
 DB 4 MELAAWCMGFLALLPFGIAGVCTGTDMKRLPASPEHLDMLRHLVYGCQVYQNL 63  
 QY 61 ELTYLPTNASLFLQDIOEVGYLHANOVRVPLQRLIRVGTQLFEDNALAVLNDG 120  
 DB 64 ELTYVPANASLFLQDIOEVGYLHANOVRVPLQRLIRVGTQLFEDNALAVLNDK 123  
 QY 121 DPLNNTTPTVT-GASPGGLRELQRLSTELIKGGLVLIQRPOLCYODTIMKDI FHKNNQL 179  
 DB 124 DPGDNVAASPRGTPGRLRELQRLSTELIKGGLVLIQRPOLCYODTIMKDI FHKNNQL 183  
 QY 180 ALLIOTNSRACHPCSPMKSGRCWGESSEDCQSLTRIVCAGGACRCKGRLPTDCCHQ 239  
 DB 184 APVDIDTNRBRAPCPACPKCKMHCWGESSEDCQSLTRIVCAGGACRCKGRLPTDCCHQ 243  
 QY 240 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNREGRTFGASCVTAC 299  
 DB 244 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNREGRTFGASCVTTC 303  
 QY 300 PNYVISTDVSGCTLVCPPLHNOEYTAEDGTORCEKSKPCARVCGYIGMOYIKANSKFI 359  
 DB 304 PNYVISTEVGSCCTLVCPPNQOETVADGTORCEKSKPCARVCGYIGMOYIKANSKFI 363  
 QY 360 TELEFAGCKKIFGSLFLPESFDGPDASNTAPLQPOLOVFEETLEITGLYISAMPDL 419  
 DB 364 NVOEFQCKKIFGSLFLPESFDGPDSSGAPLRPOLOVFEETLEITGLYISAMPDL 423  
 QY 420 PDLVFQNTQVIRGRILHNGAVSLTQIGISWLGRLSRELGSGLALIHNTLFCVHT 479  
 DB 424 RDLVFQNTQVIRGRILHNGAVSLTQIGISWLGRLSRELGSGLALIHNTLFCVHT 483  
 QY 480 VPNDQLFERNHQLLHTANRDE-CVGBGLACHOCAGHCHGSPRTCCVNSQTLRQ 538  
 DB 484 VPNDQLFERNHQLLHTANRDE-CVGBGLACHOCAGHCHGSPRTCCVNSQTLRQ 543  
 QY 539 ECVBGRVLQGLPREYVNAHCLPCHPECCOPNGSVTCGPEADQCVAAHYKDPFCVA 598  
 DB 544 ECVBGRVWKGLEPREYVNDKCLPCHPECCOPNGSVTCGPEADQCVAAHYKDPFCVA 603  
 QY 599 RCSSGVKPDLSYPIKPFPEBACQPCPINTCHSCVDLDDKCCPAEQASPLTISVAV 658  
 DB 604 RCSSGVKPDLSYPIKPFPEBACQPCPINTCHSCVDLDDKCCPAEQASPLTISVAV 663  
 QY 659 VGILLVVLGVFGIL-----FNNFTVSFMLRVKVSASHLEPTSPGAMPNQOMKI 712  
 DB 664 EGVLLFLILVVVVGILIKRRQKIRKYTRALLQETEL-----VEPLTPSGAMPNQOMKI 719  
 QY 713 LKTELKRVKVLCSGAFGYKGIWIPDGENVKIPVAILKRLRENTSPKANKETLDEAYM 772  
 DB 720 LKTELKRVKVLCSGAFGYKGIWIPDGENVKIPVAILKRLRENTSPKANKETLDEAYM 779  
 QY 773 AGVSPVSRLLIGICLTSTVQVLTOLMPYGCCLDHYRENGRGLGSDDLNMCQILKGS 832  
 DB 780 AGVSPVSRLLIGICLTSTVQVLTOLMPYGCCLDHYRENGRGLGSDDLNMCQILKGS 839  
 QY 833 YLEDAVRLVARDLAAANVLVKSBNHVKITDFGLARLLIDETEVHADGKVPKIMWALS 892  
 DB 840 YLEDAVRLVARDLAAANVLVKSBNHVKITDFGLARLLIDETEVHADGKVPKIMWALS 899  
 QY 893 LRERFTHQSDVMSVGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMI 952

DB 900 LRERFTHQSDVMSVGVTVWELMTFGAKPYDGI PAREIPDLLEKGERLPQPPICTIDVYMI 959  
 QY 953 MYKCMIDSERPPRELIVSEFSMARPPQRFVYVQNDLGPAPLSTFYRSLLEDDM 1012  
 DB 960 MYKCMIDSERPPRELIVSEFSMARPPQRFVYVQNDLGPAPLSTFYRSLLEDDM 1019  
 QY 1013 GDLYDAEYLVPOQGFPCDDPAPAGGVHHRHSSSTRSGCDLTLGLESEEEAPRSP 1072  
 DB 1020 GDLYDAEYLVPOQGFPCDDPAPAGGVHHRHSSSTRSGCDLTLGLESEEEAPRSP 1079  
 QY 1073 LAPSEGASDVFDGDLGKAAGLQSLFTHDPSELQRYSEDPYVPLPSETGYVAPLTCS 1132  
 DB 1080 LAPSEGASDVFDGDLGKAAGLQSLFTHDPSELQRYSEDPYVPLPSETGYVAPLTCS 1139  
 QY 1133 PQPEYVNOCDVPPPPSPREGPLPAPAPAGATLBRAKTLSPGKGVVDVFAFGAVENP 1192  
 DB 1140 PQPEYVNOCDVPPPPSPREGPLPAPAPAGATLBRAKTLSPGKGVVDVFAFGAVENP 1199  
 QY 1193 EYLPFGGAPAPPPPPSPREGPLPAPAPAGATLBRAKTLSPGKGVVDVFAFGAVENP 1252  
 DB 1200 EYLPFGGAPAPPPPPSPREGPLPAPAPAGATLBRAKTLSPGKGVVDVFAFGAVENP 1259  
 QY 1253 V 1253  
 DB 1260 V 1260

## RESULT 3

148161  
 P-185 precursor - golden hamster  
 C:Species: Mesocricetus auratus (golden hamster)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Jun-1999  
 C:Accession: 148161  
 R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nago, M.; Arai, M.; Yamazaki, Y.; Ishi  
 Gene 140, 251-255, 1994  
 A>Title: Cloning and activation of the Syrian hamster neu proto-oncogene.  
 A:Reference number: 148161; MUID:94193007; PMID:7908275  
 A:Accession: 148161  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1254 <RES>  
 A:Cross-references: GB:D16295; NID:9493236; PID:BA03801.1; PID:9747595  
 A:Gene: neu  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP  
 F:718-983/Domain: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.1%; Score 5801.5; DB 2; Length 1254;  
 Best Local Similarity 84.8%; Pred. No. 1, 2e-231;  
 Matches 1068; Conservative 66; Mismatches 114; Indels 11; Gaps 3;

QY 1 MELALCRWGLLALLPFGAASVCTGTDMKRLPASPEHLDMLRHLVYGCQVYQNL 60  
 DB 1 MELAAWCMGFLALLPFGIAGVCTGTDMKRLPASPEHLDMLRHLVYGCQVYQNL 60  
 QY 61 ELTYLPTNASLFLQDIOEVGYLHANOVRVPLQRLIRVGTQLFEDNALAVLNDG 120  
 DB 64 ELTYVPANASLFLQDIOEVGYLHANOVRVPLQRLIRVGTQLFEDNALAVLNDK 123  
 QY 121 DPLNNTTPTVT-GASPGGLRELQRLSTELIKGGLVLIQRPOLCYODTIMKDI FHKNNQL 179  
 DB 124 DPGDNVAASPRGTPGRLRELQRLSTELIKGGLVLIQRPOLCYODTIMKDI FHKNNQL 183  
 QY 180 ALLIOTNSRACHPCSPMKSGRCWGESSEDCQSLTRIVCAGGACRCKGRLPTDCCHQ 239  
 DB 184 APVDIDTNRBRAPCPACPKCKMHCWGESSEDCQSLTRIVCAGGACRCKGRLPTDCCHQ 243  
 QY 240 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNREGRTFGASCVTAC 299  
 DB 244 CAAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNREGRTFGASCVTTC 303

QY 301 YNVLSTDVSGCTLVCPHLNQEVTAEADGTORCEKCSQCARVCYGLCMQYIKANSKFTGTT 360  
 Db 301 YNVLSTEVGSCCTLVCPHLNQEVTAEADGTORCEKCSQCARVCYGLCMQYIKANSKFTGTT 360  
 QY 361 ELFEAGCKTIFGSLAFIPESFDDDPASNTAPLOPELOVETETETGYLYISAMPDSIP 420  
 Db 361 IQEPAGCKTIFGSLAFIPESFDDDPASNTAPLOPELOVETETETGYLYISAMPDSIP 420  
 QY 421 DLSEFQNLQYIRGRILNGAYSLTLQGLISWLGSLRELGSGLAIHHNTHLCFVHTV 480  
 Db 421 DLSEFQNLQYIRGRILNGAYSLTLQGLISWLGSLRELGSGLAIHHNTHLCFVHTV 480  
 QY 481 PMOOLFNPQALHTNRPEDVEGEGIALCHOLCARGHWGSPGPOCVNCSQFLAGQEC 540  
 Db 481 PMOOLFNPQALHTNRPEDVEGEGIALCHOLCARGHWGSPGPOCVNCSQFLAGQEC 540  
 QY 541 VEEGRVLDQGLPREVYNARHCLPCHPECOPONGSVTCFGRPADOCVACAHYKDPFCVARC 600  
 Db 541 VKEGRVWKLPREVYNARHCLPCHPECOPONGSVTCFGRPADOCVACAHYKDPFCVARC 600  
 QY 601 PSQVPLSTYPIWKFPEDEGACOPCINCTHSCVDLDEKCAEGASFLTSVNAV 660  
 Db 601 PSQVPLSTYPIWKFPEDEGACOPCINCTHSCVDLDEKCAEGASFLTSVNAV 660  
 QY 661 ILVVLGVVFGIL-----FNNFTVSFVLKVPKVSASHLEPLTPSGAMPNOQMRILK 714  
 Db 661 ILVVLGVVFGIL-----FNNFTVSFVLKVPKVSASHLEPLTPSGAMPNOQMRILK 714  
 QY 715 ETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIVLEBNISPKANKILDEAYVMAG 774  
 Db 715 ETELRKVKVLGSGAFGVYKGIWIPDGENVKIPVAIVLEBNISPKANKILDEAYVMAG 774  
 QY 775 VGSFYVSRLLIGICTSTVQVLTQMPYGLDDEHRENGRIGSODLLNMCQIAKMSYL 834  
 Db 775 LGSFYVSRLLIGICTSTVQVLTQMPYGLDDEHRENGRIGSODLLNMCQIAKMSYL 834  
 QY 835 EDVELVHRDLAARNVLYKSPNVKITDFFGLARLLIDETEHADGKVPKIMMALESILR 894  
 Db 835 EDVELVHRDLAARNVLYKSPNVKITDFFGLARLLIDETEHADGKVPKIMMALESILR 894  
 QY 895 RRFTHQSDVMSYGTWELMTFEGAKPDGIPAREIPDLEKGEELPOPPICCTIDVYIMV 954  
 Db 895 RRFTHQSDVMSYGTWELMTFEGAKPDGIPAREIPDLEKGEELPOPPICCTIDVYIMV 954  
 QY 955 KCMWIDSECRPRFELVSEFSRMARDPQRFVIONEDLGPASPLDSTFYSLLEDDMGD 1014  
 Db 955 KCMWIDSECRPRFELVSEFSRMARDPQRFVIONEDLGPASPLDSTFYSLLEDDMGD 1014  
 QY 1015 LVDAEYLVPOQGFPCPPAPAGAGMWHRRSSSTSGGDLTLGLPESEEARPSPLA 1074  
 Db 1015 LVDAEYLVPOQGFPCPPAPAGAGMWHRRSSSTSGGDLTLGLPESEEARPSPLA 1074  
 QY 1075 PSEGAQSDVFDGDLGMAKGLQSLPTHDPSPLOKYSHPVPLPSETDGVYADLTSPQ 1134  
 Db 1075 PSEGAQSDVFDGDLGMAKGLQSLPTHDPSPLOKYSHPVPLPSETDGVYADLTSPQ 1134  
 QY 1135 PEYVNPQVPRPQPSRPGPILPAPAPATLERAKTISPGKNGVYKVFAPAGVENPEY 1194  
 Db 1135 PEYVNPQVPRPQPSRPGPILPAPAPATLERAKTISPGKNGVYKVFAPAGVENPEY 1194  
 QY 1195 LTPQGAAPQPPAPAFSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYLDLPV 1254  
 Db 1195 LTPQGAAPQPPAPAFSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYLDLPV 1254

## RESULT 4

GGHUE  
 epidermal growth factor receptor precursor - human  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB  
 C:Species: Homo sapiens (man)  
 C:Date: 15-Nov-1984 #sequence revision 27-Nov-1985 #text change 11-Jun-1999  
 C:Accession: A00641; A25772; S30024; A38672; A00642; A43615; A23062; A05281; A60143; A33

R.Ullrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee,  
 K.G. P.H.  
 Nature 309, 418-425, 1984  
 A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expressi  
 A:Reference number: A00641; MUID:84219729; PMID:6328312  
 A:Accession: A00641  
 A:Molecule type: mRNA  
 A:Residues: 1-1210 <UTL>  
 A:Cross-references: EMBL:X00588; NID:G31113; PIDN:CAA25240.1; PID:G757924  
 A:Note: The authors translated the codon AAG for residue 540 as Asn  
 R.Ishii, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985  
 A:Title: Characterization and sequence of the promoter region of the human epiderma  
 A:Reference number: A25772; MUID:85270438; PMID:2991899  
 A:Accession: A25772  
 A:Status: translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-29 <ISH>  
 A:Cross-references: GB:M11234; NID:G181981; PIDN:AAA52370.1; PID:G553272  
 R.Haley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ullrich, A.; Waterfield, M.  
 Oncogene Res. 1, 375-396, 1987  
 A:Title: The human EGF receptor gene: structure of the 110 kb locus and identificat  
 A:Reference number: S30024; MUID:88217333; PMID:3329716  
 A:Accession: S30024  
 A:Molecule type: DNA  
 A:Residues: 1-29 <HML>  
 A:Cross-references: EMBL:X06370; NID:G31118; PIDN:CAA29668.1; PID:G31119  
 R.Haley, J.D.; Waterfield, M.D.  
 J. Biol. Chem. 266, 1746-1753, 1991  
 A:Title: Contributory effects of de Novo transcription and premature transcript ter  
 A:Reference number: A38672; MUID:91107677; PMID:1988448  
 A:Accession: A38672  
 A:Molecule type: DNA  
 A:Residues: 1-29 <HML>  
 A:Cross-references: GB:M38425; NID:G181977; PIDN:AAA63171.1; PID:G553271  
 A:Experimental source: carcinoma cell line A431-7  
 R.Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.  
 Nature 309, 806-810, 1984  
 A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of  
 A:Reference number: A00642; MUID:84245835; PMID:6330563  
 A:Accession: A00642  
 A:Molecule type: mRNA  
 A:Residues: 150-187, 'KSVIOAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 30  
 A:Residues: 713-964 <LIN>  
 A:Experimental source: epidermoid carcinoma cell line A431  
 R.Stamen, F.A.; Gope, M.L.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.  
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984  
 A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplifica  
 A:Reference number: A43615; MUID:84196372; PMID:6326261  
 A:Accession: A43615  
 A:Molecule type: mRNA  
 A:Residues: 713-964 <LIN>  
 A:Experimental source: epidermoid carcinoma cell line A431  
 R.Stamen, F.A.; Gope, M.L.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.  
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984  
 A:Reference number: A23062; MUID:85046483; PMID:6093780  
 A:Accession: A23062  
 A:Molecule type: protein  
 A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>  
 R.Russo, M.W.; Lukae, T.J.; Cohen, S.; Staros, J.V.  
 J. Biol. Chem. 260, 5205-5208, 1985  
 A:Title: Identification of residues in the nucleotide binding site of the epidermal  
 A:Reference number: A60143; MUID:85182650; PMID:2985580  
 A:Accession: A60143  
 A:Molecule type: protein  
 A:Residues: 740-744, 'X', 746-747 <RUS>  
 R.Moczkoewski, B.; Mosig, G.; Cohen, S.



Nature 309, 270-273, 1984  
 A>Title: ATP-stimulated interaction between epidermal growth factor receptor and superco  
 A'Reference number: A38023; MUID:84191554; PMID:6325948  
 A'Contents: annotation; receptor activity  
 A'Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA  
 R.Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.  
 Cell 59, 33-43, 1989  
 A>Title: Functional independence of the epidermal growth factor receptor from a domain  
 A'Reference number: A3331; MUID:90003233; PMID:2790960  
 A'Contents: annotation; internalization signal  
 C'Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor c  
 C'Genetics:  
 A'Gene: GDB:EGFR  
 A'Cross-references: GDB:120610; OMIM:131550  
 A'Map position: 7p12.3-7p12.1  
 C'Superfamily: epidermal growth factor receptor; protein kinase homology  
 C'Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 F1-24/Domain: signal sequence #status predicted <SIG>  
 F1-25-1210/Product: EGF receptor #status predicted <EXT>  
 F1-25-645/Domain: extracellular #status predicted <EXT>  
 F1-75-100/Domain: EGF receptor extracellular domain repeat <EE1>  
 F1-390-600/Domain: EGF receptor extracellular domain repeat <EE2>  
 F1-646-668/Domain: transmembrane #status predicted <TM>  
 F1-669-1210/Domain: intracellular #status predicted <INT>  
 F1-10-975/Domain: protein kinase homology <KIN>  
 F1-718-726/Region: protein kinase ATP-binding motif  
 F1-999-1046/Region: coated-pit mediated internalization signal  
 F1-1047-1210/Region: inhibitory  
 F1-128-175-355-413-444-528-603/Binding site: carbohydrate (asn) (covalent) #status predic  
 F1-745/Active site: Lys #status experimental

Query Match 45.0%; Score 3069; DB 1; Length 1210;  
 Best Local Similarity 49.1%; Pred. No. 3e-119;  
 Matches 621; Conservative 175; Mismatches 361; Indels 108; Gaps 24;

```

QY 11 LLLALLPAGA--STVCTGCTPMKRLPASPETHLDMHLVYOGCVVQGNELTYLPTN 68
DB 14 LLLALCPASALBEKKVCCGTGSKNLTQLTFEDHFLSDRMNCEVVLGNLEITYVQN 73
QY 69 ASLSFLQDIQEVQVYLHNQYRQVFLRLIVRGTLFEDNYALVLDNGDPLNNTTP 128
DB 74 YLSFLKTIQEVAGVYLILNTERIPLENLQIRGNMYENSVALAVLSND----- 126
QY 129 VVGASPGRLREIQLRETELKGVILIGNPOLCYQDTLWMDIFHKNQNLALLIDTNR 188
DB 127 ---ANKTGLKEIPMRNLQILHGAVFSPNNPALCNVESTIQMDIVSSDLSNMSDPQNH 183
QY 189 SRACHPSCPMKGRSCWGSSESDQSLRTVCAAGCA-RCKGRLPTCCHEQCAAGCTGP 247
DB 184 LQSCQKCDSPCGSCWGAEEHCQLTKILCAQCSGRGRGKSPEDCCNCAAGCTGP 243
QY 248 KASDCLACLFPHNSGICELHCPALVTYNNTDFESMNPBGRYTFGASCTYACRYVLTSD 307
DB 244 RSDCLVCKRFDEATCKTCTCPMLYNFTYQMDVNPFGKTSFGATCYKCKPRYVYVTD 303
QY 308 VGSCTLVCPHLNQEVLAEDGTQCEKSCSKPCARVCGYGLQMYIKANSKEITGTELE-FAG 366
DB 304 HSCVVAACGADSEV-EDQGVKCKCKCEGPCRVCGIGIERK-DLSLNNNTNKHFGN 361
QY 367 CKKIFGSLAFPSFSGDPAANTAPRLQVQLQVFTLEITGYLISAMPDLPDLSTVQ 426
DB 362 CTISGDLHLPLPAFGSDFTHTPPLDPQLILKVKKITGFLILQAMPERTDLHAE 421
QY 427 NLOVIRGRIHNGAVSLTQGLISVWLGRLSRELSGALIHNNTHLCFVHTVPDQLF 486
DB 422 NLEIIRGRKQHQGFSLAVSLNITSLGRSLKEISDGVIIISGNGLCYANTIMKKLF 481
QY 487 RNPQALLHTANPEDECVGEGLAQCLCARGHGCPGPGTQCVNCSQFLRQGECEVECRV 546
DB 482 GTSQGTXTKIISNGENSCATGQVCHALCSPEGCWPEPRDVCVSNVRSGRECVDKXL 541
QY 547 LOGLPREYVNAHCLCHEPCOPONGSVTCFGEADQCAACAHYKDPFCVAVRCSGVXP 606
  
```

```

DB 542 LEGEPREVENSECIQHPCECLPQAMNITCTGRGPDNQCIAHYIDGPHVCTCPAGW 601
QY 607 DLSMPLMKPEDEGACQPCPINTGSHCVLDLDDKCPAEARASPLTISVAVVQ---ILL 663
DB 602 ENNTL-VWKIADAGVCHLCHPNTCTCTGPGLEGGCTTNGPKP--SIATGMVALLLL 658
QY 664 VVVLGVVFGILLFNNTFV-SFWLRVPKYSASHLEPLTPSGAMPQAOAKRILKETELRKV 722
DB 659 VVALGI--GLFMRRRIHVRKRLRLRLQERELVPLTPSGEAPQALLRLIKETEFKXIK 716
QY 723 VLGGAGCTGYKGIWIDGENVKIPVATKLRNTSKAKELLDEYLVAVAGVSPVYSR 782
DB 717 VLGGAGCTGYKGIWIDGENVKIPVATKLRNTSKAKELLDEYLVAVAGVSPVYSR 776
QY 783 LLGICLSTVQLVQLMPYGLLDHVAENGRGLSDLLWCMQIAKMSYLEEDVRLVR 842
DB 777 LLGICLSTVQLIQLMPFGGLDYVAEHNDIGSQVLLMVCVQIAAGMVLLEBRLLVR 836
QY 843 DLARNVLVSPNHVKTIDFGLAALDIDETEHYADGKVPKMMALESTLRRTQSD 902
DB 837 DLARNVLVKTPOHVKTIDFGLAALDIDETEHYADGKVPKMMALESTLRRTQSD 896
QY 903 VMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPEPCTIDYVMIMVKCMIDSE 962
DB 897 VMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPEPCTIDYVMIMVKCMIDAD 956
QY 963 CRPFREIVSEFSPBARDPQRFVYIQ-NEDLGPASPLDSTFYRSLDEDDMGDLVDAEY 1021
DB 957 SRPFREILIFBSKWARDPQGVYVLIQDERMHLSPDSNRYRLMKEEDDDVDVDAEY 1016
QY 1022 LVPOQGFPCPDPAAGAGMHHRRSSSTSSGGDDLGLPSESEAPRSLASEGGS 1081
DB 1017 LIPQGF-----SSPSTSRPLSLSTSS 1042
QY 1082 DVFDGDLQMAKGLQSLPTHDPSPRLQRYSEDPVPLPSEF--DGVAFLTCSQPEEYV 1139
DB 1043 N-NSTVACIDRNLDQSPFIDESEFLQRYSDPICALTEOSIDDTFL-----VPEVIN 1094
QY 1140 QPDVPPPPSREGPLPAARAGATLEBAKTLGSKGKGVKDVAFGAVNPEYL-TPQ 1198
DB 1095 Q-SVPRKAGSVQNPVYHNOPLN-----ABSRPHYQD--PHSTAVNPEYLVNQ 1143
QY 1139 GAAPQHPHPAPFAPFNLYVQD-----DP-----PERGAPSTFGPTPAE 1242
DB 1144 -----PFCVNSTPDSPAHMAQKSHQSLNDPQYQDPFKEAKPNTIFKGS-TAE 1193
QY 1243 NPEYL 1247
DB 1194 NAEYL 1198

```

RESULT 5  
 A33183  
 epidermal growth factor receptor precursor - mouse  
 C'Species: Mus musculus (house mouse)  
 C'Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999  
 C'Accession: A53183; A43818; S24942; A28941; S43325; 149643  
 R'Location: N.C.; Philadelphia, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N  
 Genes Dev. 8, 399-413, 1994  
 A>Title: The mouse waved-2 phenotype results from a point mutation in the EGF recep  
 A'Reference number: A53183; MUID:94170986; PMID:815255  
 A'Accession: A53183  
 A'Molecule type: mRNA  
 A'Resides: 1-1210 <LUE>  
 A'Cross-references: GB:U03425  
 R'Avioli, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morze, B.  
 Oncogene 6, 673-676, 1991  
 A>Title: Comparison of EGF receptor sequences as a guide to study the ligand binding  
 A'Reference number: A43818; MUID:91232866; PMID:2030916  
 A'Accession: A43818  
 A'Molecule type: mRNA  
 A'Resides: 1-714 <AVT>  
 A'Cross-references: GB:X59598

R.Eisinger, D.P.; Serrero, G.  
 submitted to the EMBL Data Library, June 1992  
 A:Reference number: S24942  
 A:Accession: S24942  
 A:Molecule type: mRNA  
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>  
 A:Cross-references: EMBL:212608  
 R:Heiserman, G.J.; Gill, G.N.  
 J. Biol. Chem. 263, 13152-13158, 1988  
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated  
 A:Reference number: A28941, M01D:88310814; PMID:3138233  
 A:Accession: A28941  
 A:Molecule type: protein  
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,  
 R:His, M.L.; Dunn, A.R.; Alexander, W.S.  
 submitted to the EMBL Data Library, April 1994  
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor  
 A:Reference number: S45325  
 A:Accession: S45325  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1971, 'X', 973-1210 <VER>  
 A:Cross-references: EMBL:X78987, NID:9486830; PID:CA55587.1; PID:9486831  
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Day, S.K.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993  
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse  
 A:Reference number: 149643, M01D:93126380; PMID:7678348  
 A:Accession: 149643  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 12-20, 22-132 <RES>  
 A:Cross-references: GB:L06864; NID:9193001; PID:AA53029.1; PID:9567201  
 C:Genetics:  
 A:Gene: PGR  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop  
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop  
 F:1-24/Domain: signal sequence #status predicted <SIG-  
 F:648-970/Domain: transmembrane #status predicted <SIG-  
 F:712-977/Domain: protein kinase homology <KIN>  
 F:720-728/Region: protein kinase ATP-binding motif  
 F:680-695/Binding site: phosphate (Thr) (covalent) #status experimental  
 F:697-1070/Binding site: phosphate (Ser) (covalent) #status experimental  
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental  
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental  
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 44.6%; Score 3040; DB 2; Length 1210;  
 Best Local Similarity 48.9%; Pred. No. 4,7e-118;  
 Matches 622; Conservative 168; Mismatches 369; Indels 114; Gaps 25;

364 -FACCKIFGSLATLPSFDODPASNTAPLOPELOVFEETLEETGYLYISAMPDSDL 422  
 358 HFXYCTAISGLHILPAFKGDSFTIRPPDRELEIKVLEITGLLLQAMPDNDTL 417  
 423 SVFONLVRIRILHNGAVSLTLOGLSWLGSLRELSGLALIHNNHLCFVHTVPW 482  
 418 HAFENLEIRRTQHOFSLAVGLNITSGLRSLNEISGDVILISGNRLCYANTINW 477  
 483 DQLFRRPHQALLTAARPEDECVSEGLACHOLCARHGCKWGPPOCNCSFLRGSCVE 542  
 478 KKLFGTPQKTKIMNNNAEKCKAVNVHNCPLCSSEGCWPEPRDVCVQNVSGRCVE 537  
 543 ECRVLOGIPREYVNAHCLPCHEPCQPNQSVTCFGEADOCVACAYKDPFPCVACPS 602  
 538 KCNLEGEPRREFVNSQCQHPRCLQANNITGTGSPDNCLQAHYIDGPHCVKCPA 597  
 603 GVKPDLSYMPKWFPEDEGACQPCPINCTHSVDLDKCGPAEORASPLTISVAVGIL 662  
 598 GIMSENNTL-VMKYADANNVCHLCHANCYGCAGPGLQGEVWPSGKIPSLIAGIYGL 656  
 663 LVVLGVVFGILF--NNFTVSFWLVRPKVSASHLEPLTPSGAMPNQAQRILKETELR 719  
 657 LPIV-VALGIGLFRRRHIVKKTTLRLLORELEVEPLTPSGAPNQAHRILKETELR 715  
 720 KVKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKELIDEAVYMGVSPY 779  
 716 KIKVLGSGAFGTVYKGIWIPDGENVKIPVAIKVRENTSPKANKELIDEAVYMGVSPY 775  
 780 VSRLLIGLCTSTVQVQLMPYGLLDHVRNRRGLSODLLNMCQAKMSYLEVRL 839  
 776 VCRLLIGLCTSTVQVQLMPYGLLDHVRNRRGLSODLLNMCQAKMSYLEVRL 835  
 840 VARDLAARNVLYKSPNHVKTIDFGARLDIDETRYADGKVEIKMALESILRRRPTH 899  
 836 VHRDLAARNVLYKSPNHVKTIDFGARLDIDETRYADGKVEIKMALESILRRRPTH 895  
 900 QSDVSYGVTVWEIMTFGAKPYDGIIPAREIPDLLEKGRLEPQPICTIDYVIMVCKMI 959  
 896 QSDVSYGVTVWEIMTFGAKPYDGIIPAREIPDLLEKGRLEPQPICTIDYVIMVCKMI 955  
 960 DSECPREELVSEFSRMAKPPQFVVIQ-NEDGAPSLDSTFYRSLIDDDMDGLVDA 1018  
 956 DSDSPKRELILEFSQVARDPQKIVLYQSEBRNHLPSPTSNRYALMDEDEDDVDA 1015  
 1019 EBYLVQGGFFCQPPAPGAGMVHRRSSSTRSGGDLTLGLEPSEEARPSPLABSEG 1078  
 1016 DEYLIPQGGFF-----NSPST-----SRTPLSLSL 1041  
 1079 AGSDVFDGDLGMAKGLQSLPTHDSPLQYSDPTVPLPSET--DGVAAPLTCSPQPE 1136  
 1042 ATSN---NSTVACINRNGSCRKVEDAPLQYSSDPTGAVEDNIDDAFL-----PVPE 1091  
 1137 YVNPDPVAPPPSPREGPLPAAPAGATLEBAKTLSPQKGVNVXDVAFGAVENPEYL- 1195  
 1092 YVNO-SVYKRPAGSVQNVHNPQLHP-----APGRDHYQV--PHSNAVGNPEYLN 1140  
 1136 TPQGAQAQPHPPAPFPAFNDLVYWD-----DP-----PERGAPPSFCKPTP 1229  
 1141 TAQ-----PTCLSSGFNSPALWLOKSHQMSLDNPDYQDFFPKETKNGIFKG-P 1190  
 1240 TAENPEYGLDVP 1252  
 1191 TAENAEYLRVAPF 1203

RESULT 6  
 TSVCHV  
 epidermal growth factor receptor precursor - chicken  
 N:contains: protein-tyrosine kinase (EC 2.7.1.112) erbB  
 C:Species: Gallus gallus (chicken)  
 C>Date: 28-Feb-1986 #sequence\_revision 05-May-1995 #text\_change 04-Feb-2000  
 C:Accession: A27720; A00643

R/Lax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennet  
Mol. Cell. Biol. 8, 1970-1978, 1988  
A>Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou  
A:Reference number: A27720; MUID:88261272; PMID:3260329  
A:Accession: A27720  
A:Molecule type: mRNA  
A:Residues: 1-1223 <LAX>  
A:Cross-references: GB:M20386  
R/Nilen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Ciltenden, L.B.; Raines, M  
Cell 41, 719-726, 1985  
A>Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro  
A:Reference number: A00643; MUID:85228222; PMID:2988784  
A:Accession: A00643  
A:Molecule type: mRNA  
A:Residues: 585-1223 <NIL>  
A:Cross-references: GB:M10066  
C:Genetics:  
A:Gene: erbB  
C:Superfamily: epidermal growth factor receptor; protein kinase homology  
C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor  
C:Specific protein kinase  
F:1-30/Domain: signal sequence #status predicted <SIG>  
F:31-1223/Product: epidermal growth factor receptor #status predicted <MNT>  
F:31-654/Domain: extracellular #status predicted <EXT>  
F:61-307/Domain: EGF receptor extracellular domain repeat <EEL>  
F:655-677/Domain: transmembrane #status predicted <TMM>  
F:678-1223/Domain: intracellular #status predicted <INT>  
F:719-984/Domain: protein kinase homology <KIN>  
F:727-735/Region: protein kinase ATP-binding motif  
F:136-202,280,361,370,422,575,580,615,635/Binding site: carbohydrate (Thr) (covalent) #  
F:92,650/Binding site: carbohydrate (Ser) (covalent) #status predicted  
F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
F:754/Active site: Lys #status predicted  
F:1100,1183,1208/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 44.3%; Score 3017.5; DB 1; Length 1223;  
Best Local Similarity 47.9%; Pred. No. 4e-117;  
Matches 621; Conservative 171; Mismatches 358; Indels 147; Gaps 27;

QY 8 RGLGALLALPPGAA-----STQVCTGTDMKRLPASPEHLDMLHLYGCGVQVGNLE 61  
DB 13 RGAALVALLLGLGALGAVEKKYCOGTNNKLTQLGHVEDHFTSLORMYNCEVLSNLE 72  
QY 62 LTLPTNASLSFODIOEVGYLLAHNOVROYLQRLRYRGTOLEFEDYALAVLDNG 121  
DB 73 ITVEHNRDLFLKTIQEVAGYVLLANMVDVPLENQLIRGNVLDNSFALAVSNH 132  
QY 122 PLNNTPTVTGASRGURELQRLSTELKGVLIQRPOLCYQDTILMDIFHKNQAL 181  
DB 133 -MKKTQ-----GLRELPMKRLSEITLNGVAKISNNPKLCNMDVYLMNDIIDTSRK-PL 182  
QY 182 TLID-TNRSACHPCSMCKGSRCKWSSSDCOSLRTVAGGCA-RCKGRLPTDCHEO 239  
DB 183 TVLDFAANLSSCPKCHNCEDHCKWAGBQNCQTLTKVITACQCSGCRKRVSDCCHQ 242  
QY 240 CAAGCTGPKHSDCLACHFNHSGICEHCPALVTYNTDPESMNPEGRYTFGASCYTAC 299  
DB 243 CAAGCTGPRSDCLACKFRDADTCXOTCPVLVYNTTYQMVNVEGKGSFATCVRE 302  
QY 300 PNYVLTSDVGSCTLVCLHNOETAEHGTORCEKSPCARVCYGLQMYIKANSFTGI 359  
DB 303 PNYVLTSDVGSCTLVCLHNOETAEHGTORCEKSPCARVCYGLQMYIKANSFTGI 360  
QY 360 TELE-FAAGCKIFGSLAFLESPFGDPASNTAPLOPEQLOVFELEITGYLISWPPS 418  
DB 361 TNDSPKCTKINGDVSLPVALGDAFTKLPDLPRKLDVFTVEISEGFLILQAMPDN 420  
QY 419 LPDLSVQNLQVIRGLLHNGAVALTQIGIGISLGRSLREISGLALIHNTLCPVA 478  
DB 421 ATPLVAFENLEIRGTRKQGVSLAVNLIKISLGRSKESISDGIAMKKNKCYAD 480  
QY 479 TVPMDQLFRNPHQALLHTANRPEDECVGEGLAGHCACAGHCGAPGPTQCVCNSQFLRQ 538

DB 481 TMNRSLEPATOSQKTKIIONRKNKDCTADRAVDCPLCDVCGMGWGPNCSCRFSSROK 540  
QY 539 ECVESCRVLQGLPREYVNAARHCLPCHRCOPONG---SVTFGPRADOCVCAHYKDPF 595  
DB 541 ECVQKCNLLQEPREFEERDSKCLPCHSECLVQNSTAVNTTSGSPRDCMKCAHIDGH 600  
QY 596 CVARCPSPGVPDLSYMPIMKPEDEGACOPCPINCTHSCVLDLDDKCPAQRASPLTSIV 655  
DB 601 CVKAPPAVLGENDTL-VMKYADANAVQQLCHPNTROCKPGLGECF---NGSKTFSIA 656  
QY 656 SAVV-GILLVYLVGVFEILLFNNFTV-SFMLRPKPVASHLEPLTBSGANPNQOMIL 713  
DB 657 AGVAGGLLCLVYVVGIGLYLRRRIYKRTLRLLQSRLEVEPLTPSGEAPNQALHRL 716  
QY 714 KETELRYKVGSGAGFTVYKGIWIPDGENYKIPIVAIKVLRNENTSPKANKELIDEAHYMA 773  
DB 717 KETEFKRYKVLQSGAGFTVYKGIWIPDGENYKIPIVAIKVLRNENTSPKANKELIDEAHYMA 776  
QY 774 GVSPPYVRLIGICTSTVOLVTLQMPYGLLDHYRENRGRGLGSDLLNMCQIAKMSY 833  
DB 777 SYDNPHVQRLIGICTSTVOLVTLQMPYGLLDHYRENRGRGLGSDLLNMCQIAKMSY 836  
QY 834 LEDVRLVNRDLAARVVLKSPNHYKITDPGLARLLDIDETETHAOGKVPKIMMALESL 893  
DB 837 LEERLVRDLAARVVLKSPNHYKITDPGLARLLDIDETETHAOGKVPKIMMALESL 896  
QY 894 RRRPTHSDWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIM 953  
DB 897 HRIYHQSDWSYGVTVWELMTFGSKPYDGIPIAREIPDLLEKGERLPQPICTIDVYIM 956  
QY 954 VKCMWIDECRRFRELVSFERSRMARDQRFVILQ-NDLGPASLDSITFRSLLEDDM 1012  
DB 957 VKCMWIDECRRFRELVSFERSRMARDQRFVILQ-NDLGPASLDSITFRSLLEDDM 1016  
QY 1013 GDLVAFEEVLPQGFPCPDPAAGAGVHHRHSSSTRSGGDTLLGLEPSEERAPSP 1072  
DB 1017 EDIVYADELVHQQFF-----NSPST-----SSTP 1042  
QY 1073 L-----APSEGASGVFPGDLAGAKGLQSLPTHDPBLQRYSDPTVPLPSET--DGY 1125  
DB 1043 LLSLSLANSNNSATNCID-----RNGOGHPREDEDFVQRYSDPTGPNFLESIDGIF 1094  
QY 1126 VAPLFCSPQPEVYVNDPVRPOPSPRDEBPLPAARAGATLEBAKTLSPGKGVADV-- 1183  
DB 1095 L-----PAPEYVQ--LMPKPS-----TAMQONTYNNISLT 1125  
QY 1184 -----AFGAVENPEVLTPOGGAPOHPPPAPSPAFDNLVYWDQ----- 1223  
DB 1126 AISKLPMDSRVYNSHSTAVDNEPL-----NTNQSPLAKTVFESSPYWIOSGNHOIN 1177  
QY 1224 -DPE-----RGAPSTFKCTPTAENPEVGLDVP 1252  
DB 1178 LNDPDIQODFLPNETKPNGLKVPAPENPEVLRVAP 1214

RESULT 7  
A47253  
epidermal growth factor receptor, HER4 - human  
C:Species: Homo sapiens (man)  
C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999  
C:Accession: A47253  
R:Blowman, G.D.; Culnan, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L  
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993  
A>Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide  
A:Reference number: A47253; MUID:93189574; PMID:8383326  
A:Accession: A47253  
A:Status: preliminary; not compared with conceptual translation  
A:Molecule type: nucleic acid  
A:Residues: 1-1308 <PIO>  
A:Cross-references: GB:I07868; NID:9337359; PIDN:AAB59446.1; PID:9337360  
A>Note: sequence extracted from NCBI backbone (NCBI:126842)  
C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP, growth factor receptor  
 F:116-981/Domain: protein kinase homology <Kin>  
 F:124-732/Region: protein kinase ATP-binding motif

Query Match 42.1% Score 2908.5; DB 2; Length 1308;  
 Best Local Similarity 44.1%; Pred. No. 13e-112;  
 Matches 600; Conservative 182; Mismatches 384; Indels 185; Gaps 28;

9 WGLLLALPPGAA-----STQVCTGDMKRLRPASPTLMDLRLHLYOGQVVOGNIETLY 64  
 8 WWSVLLVAGTVQPSDSQSVACAGTENKLSLSLDEQQYRALRKYYENCEVWVGNIETLS 67  
 QY LPPNNAISFLQDIOEOGYVLIANNOVROVLORLIVRGTOLEFENYVALVYDNDDELN 124  
 65 LPPNNAISFLQDIOEOGYVLIANNOVROVLORLIVRGTOLEFENYVALVYDNDDELN 124  
 Db IERNRDLSELRSEVRETVGVVVALNPFYVLENLKIIRKTYEPRYALATLRYKRG 127  
 QY 125 NTPVTVGASPGGLRELOLRSLTEILKGVLIQNPOLCYODTILMKDIFHKONQALTLI 184  
 128 NF-----GLQELGLKNLITELINGVYVDONKFLCYADTIHMODIVNMPVSNLLTV 178  
 QY 185 DTRRSRACHPCSPCKSGRCMGSSSEDCSLTRTVACAGC-AECCKPLPTDCCHCCAG 243  
 Db 179 STNGSGGGRCHKSCGTG-RCWGPTEHNCQTLNRTVACBQDCGCGYGVYSDCHRCAG 237  
 QY 244 CTGPRHSDCLACLFHNSGICELHCPALVTYNTDTESEMPREGRYTFGASCVTAQPYN 303  
 Db 238 CSGPKOTDCAFCAMFNFGSACVTCQCPQTFVYNPTTFLNHNFAKTYGAFVCKKCPHNF 297  
 QY 304 LSTVSGCTLVCPILHNOEVAEDGTQCEKSKPCARVCYGLGMQYIKANSKFTIGTELE 363  
 298 V-VDSSCVACPCSKMEV-EENGIKWCKCKCTDIPCACGIGIGSLSMQTVSSNIDK 355  
 QY 364 FAGCKIFGSLAFPESEDDPASNATAPLOPEQLOVETLEETIGYLYISAMPDLPJLS 423  
 Db 356 FINCKTNGNLIETVTHGIDPRYALAIIDPEKLNVRTRYREITGFINLSQWPNMTDS 415  
 QY 424 VFQNLQVIRGRIHNGAYSLTQGLGISWLGSLRLSGGLAIHHNHTLCHVHYVPM 483  
 Db 416 VFSNLTVIGGRVLTYSGLSLILKQGITSLQFOSLKEISAGNIYITDNSLVCYHTINMT 475  
 QY 484 QLFENPHQALHTNRPEDCEVGEGLACHOLCAAGHGMGPGTCCVCSQFLLGOECVEE 543  
 Db 476 TLFSTINORIYIRNRKAENCTAEGMCMNHCSSDGMGPGPOCLSCRFSGRITCS 535  
 QY 544 CRVLQGLPREVYNAHCLPCHPECQP-QNSVSTFGSEADQVCAAHYKDPPECVACPS 602  
 Db 536 CNLYDEPERFENSGICVECDPQCEKMEGDLTCHGPGPNCTKSHFKGPGNCEKCPD 595  
 QY 603 GVRPDLGMPYWKPEDEGACQPCPINCTHSCVDLDDKGC-----PAEQRASPL 651  
 Db 596 GLQGANST--LFXADPDRCHPCHPNCCTQGCNCPSTHDCIYFMTGHTSLPQAR-TPL 652  
 QY 652 TSISAVV-GILLVVVLGVFGLIFNNFTVSWLRVYKVSASH-----LEBLPSGA 703  
 Db 653 --IAAGVIGLFIIVGLFPAVYVRK-----SIKKKALRRFETELVEPLTSPSGT 703  
 QY 704 MPNOQKRLKETELRKRYLGSAGFGVYKGIWIPGNGKIKVAILKULRENTSPRANK 763  
 Db 704 APNOQKRLKETELRKRYLGSAGFGVYKGIWIPGNETYKILVAILKINETTPRANV 763  
 QY 764 EILDAAVYAGVSPVYRLLIGLICLTSTVQLVTOLMPYGCGLDHNRENRGLSGODLLN 823  
 Db 764 EFMDEALIMAGMDPHLVRLGLVCSLPTIQLVTOLMPHGLLEVEHKNIGSOLLN 823  
 QY 824 CMQIAKMSYLEDVRLVHRLAARNVLYKSPNHKITYDFGLARLLIDETRYHADGKVP 883  
 Db 824 CVOIAKMSYLEDVRLVHRLAARNVLYKSPNHKITYDFGLARLLIDETRYHADGKVP 883  
 QY 884 IKMMALESILRRRFTHQSDVSYGYVWELMTFGAKPYDGIPIAREIPDLLEKGERLQPP 943  
 Db 884 IKMMALESILRRRFTHQSDVSYGYVWELMTFGAKPYDGIPIAREIPDLLEKGERLQPP 943  
 QY 944 ICTIDVYIMVCKMIDSECRPRFELVSEFSKARDQRFVVIQNEB-LGPASPLDSTF 1002

Db 944 ICTIDVYIMVCKMIDSECRPRFELVSEFSKARDQRFVVIQNEB-LGPASPLDSTF 1003  
 QY 1003 YRSLEDDKGDVDAEYLYVPOGFPDPAAGAMHRRSSSTRSGGLTLCLE 1062  
 Db 1004 FQNLDEBDEDDMDAEYLYV-CAFNIIPP-----IYSSRAIDSR-----EIGHS 1051  
 QY 1063 PSEBAPPS-----PLAP-SEGAGSDVFDGLGMGA 1093  
 Db 1052 PPAVTPMSGQFYRRDGGFAEAGVSPYPAFTSTIPAPVAGATETIIDDSCNCTL 1111  
 QY 1094 KGLSLPTHPDSPLORYSEDEPTVLP-S-----ETDGYVAPLTCSPOPEYVNOBDVAPQ 1146  
 Db 1112 RKPVAHVQEDSSTQRYGADPTVAPERSPRGLEDDEGYMTPMRDKPQOYLYNVE----- 1167  
 QY 1147 PPSREGLPAPAPAGATLEBAKTLSPKKNVVDVFAFGAVENPEYLTPOGGAAPQH 1206  
 Db 1168 -----ENPFVSR-----KNNDLQ-----ALDNPEYHANSNG----- 1194  
 QY 1207 PPPA-----FSPAFLNLYYMDQDPPERGA-- 1230  
 Db 1195 PPXADEVYNEPLTYNTFANTLGAKEYLKNILSMPEKAKKAFDPDYMHSLPFRSTLQ 1254  
 QY 1231 PPSTRKTP-----AENPEYL 1247  
 Db 1255 HPDYLOEYSTKYFYKQNGRIRPIVANEPEYL 1285

RESULT 8  
 S06142  
 protein-tyrosine kinase (BC 2.7.1.112) mrk-y precursor - southern platyfish  
 N:Alternate names: epidermal growth factor receptor homology; kinase-related transfo  
 C:Species: Xiphophorus maculatus (southern platyfish)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Feb-2000  
 R:Accession: S06142; S13809  
 R:Hitbrodt: J; Adam, D.; Malleschek, B.; Maeueter, W.; Raulf, F.; Telling, A.; Ro  
 Nature 341, 415-421, 1989  
 A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing T  
 A:Reference number: S06142; MUID:90015140; PMID:2797166  
 A:Accession: S06142  
 A:Molecule type: DNA  
 A:Residues: 1-1166 <MT>  
 A:Cross-references: EMBL:X16891; NID:965220; PID:CAA34770.1; PID:9652291  
 R:Adam, D.; Maeueter, W.; Scharf, M.  
 Oncogene 6, 73-80, 1991  
 A:Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xipho  
 A:Reference number: S13807; MUID:91125882; PMID:1846957  
 A:Accession: S13809  
 A:Status: preliminary; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>  
 A:Cross-references: EMBL:X56319; NID:965284; PIDN:CAA39763.1; PID:965285  
 C:Genetics:  
 A:Gene: mrk  
 A:Map position: Y  
 A:Insertions: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP, growth factor receptor; phosphotransferase; transmembrane protein;  
 F:125/Domain: signal sequence #status predicted <Sig>  
 F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MT>  
 F:707-972/Domain: protein kinase homology <Kin>  
 F:715-723/Region: protein kinase ATP-binding motif

Query Match 38.1% Score 2598; DB 1; Length 1166;  
 Best Local Similarity 44.1%; Pred. No. 7e-100;  
 Matches 561; Conservative 168; Mismatches 394; Indels 148; Gaps 31;

QY 4 AALCRWGLLALPPGAAST-----QVCTGDMKRLRPASPTLMDLRLHLYOGQVVOGNI 59  
 Db 8 AALLQ--LLVLSISRCSTDPDRKVCQSTSNQMT--LDNHYLKMKKMYSGCVVLEEN 62  
 QY 60 LETTYPTNASTFLQDIOEOGYVLIANNOVROVLORLIVRGTOLEFENYVALVYDND 119

```

Db      63 LEITYQENODLSFLQSIQEVGVYLIANNEVSTIPLVNLRLIRGQNLTEGNTLLVMSN 122
Qy      120 GBLPNTTIVTASPCGLRELOLRSLTEILKGVLIQRNPOLCYODTIMKDIHFKNOL 179
Db      123 YOK-NBSSP--DYVQGLKQLOLSNLTSLGKVNKSHNPLCNVETIMWIDVTISP 179
Qy      180 ATLIDTNSRACHPCSPMKSGRCWSESSEDCOSLTRTVCAAGC-ARCKGLPDPDCHE 238
Db      180 TNNLIPHAFERQCKCDHGCNVGSCVAPRGHCQKFTKLCAEQCRRCRGKPIIDCNE 239
Qy      239 QCAAGCTGPKHSPCLLAFHNSGICELHCPALVYNTDTPFSMPREBRTTFGASCYTA 298
Db      240 HCGAGCTGPRATDCLACRDNFDGTCDCPEPKYDIVSHOVNPNKTYFGAACVBE 299
Qy      299 CPVNYLSTVGSCTVCPHNOEVTADGTQCEKSKPCARCVGLQMGVYIKANSKFIG 358
Db      300 CPENYVTE-GACVRCSSAGMLEVD-ENKRSCKPCDGVCPVCCSIGLSL-SNTIAN 356
Qy      359 ITFL-EPAGCKTIFGSLAPLPSFDGDPASNTAPLOPELOVFEETLEITGYLYISAMPD 417
Db      357 STNIRSFNCTKINGDIIINRNSFEEDPHYKIGTMDPEHMLNLTVKEITGYLYIMMPE 416
Qy      418 SLDPDLSVFNLOVYTRIRILHNGAYS-LTQGLGISWLGRLRELSGALHHTHLCF 476
Db      417 NMTSLVFQNLREIRGRTFSRGSFVYVQVRLHQLGRLSKVAGNVILKNTQLRY 476
Qy      477 VHTVPDQLFRNHQALLHTANPEDECVGEGLAGHQLCARHCWPGPFGTCVNSQFLRY 536
Db      477 ANIMRRLFRSEDSQIEVDART-----ENQTCNNESEDCGM-PCGTWVCSCILHVR 528
Qy      537 GQCEVECEVLOGLPREVYANARHCLPHECCQPNQSNVCFGEADQCAVCAHYDPRPC 596
Db      529 GGRCVASCNLQGEPRAOVDGRCVCHOECLVQDLSLTCYGGPANCKSAHFQDGPQC 588
Qy      597 VACPGGVPRDLSYMPIMKFPDEGACOPCINCTSCVDLDDKGPACQASPLTISVS 656
Db      589 IPRCPHIGLDGDTL-IMKXADKMGCCQCHQCTGCGGPGSGRCGP-IYSHSLAVG 646
Qy      657 AVVGLLVVVLGVVFGIL-----ITNNFYSEFLKRPKXSASLLEUTSSGAMPNOAKR 711
Db      647 LVSGLLITVIALVLLRRRRIRKRRIRKRRIRKRRIRKRRIRKRRIRKRRIRKRRIR 702
Qy      712 ILKETELRKVKVLSGAFGTVYKGIWIPGENVKIPIVAILKRENTSPRANKELIDEAY 771
Db      703 ILKETELFKDRVLGSAFGTVYKGLMNPGENIRIPVAILKRENTSPRANKELIDEAY 762
Qy      772 MAGVSPYVSRLLIGLCTSTVQVLTQLMYGCILDHVRENRGLSGQDILLMCMQIAKGM 831
Db      763 MASVDHPHVCRLIGLITSAVOLVTQLMYGCILDHVRENRGLSGQDILLMCMQIAKGM 822
Qy      832 SYLEDVRLVHRDLAANVIVKSPNHVITDFGLARLDIDETFYHADGGKVPILKMALES 891
Db      823 NYIEERHVLHRDLAANVILKNNHVKITDFGLSKLITDADEKEYQADGKVPILKMALES 882
Qy      892 ILRRRTTHOSDWSYGVYWEMLTFSKAPYDGIIPAEIPDLLEKGRSLPQPICTIDVYM 951
Db      883 ILCMTYTHOSDWSYGVYWEMLTFSKAPYDGIIPAEIPDLLEKGRSLPQPICTIDVYM 942
Qy      952 IMVKCMWIDSECPRELVSEFSRMAPDPORFVVIQNDLGPASLSTFRRSLIEDD 1011
Db      943 IILKCMWIDSSFRPRELVSEFSRMAPDPORFVVIQNDLGPASLSTFRRSLIEDD 999
Qy      1012 MEDVLAEEYLVQGFCEPDPAAGGVVHHRRSSSTRSGGDLTLGLBSEEBAPRS 1071
Db      1000 --DVVDADAEYLIPYKRI-----NRQGS-----E 1020
Qy      1072 PLABSEGASDVVDGLGMAKGLQSLPTHDSPLQRYSEDPVY-PLPSETDGYAVALT 1130
Db      1021 PCLPPTPTGH-----PYRENSITLRNISPDTQALKEKDDGH----- 1055
Qy      1131 CSQPEYVYVQDPVRRP-----PSPRE-----GPLP--AARPAAGTLERAKTLSPKNG 1177
Db      1056 -----EYVNOFGSETSSRLSDIYNFYEDLTDCMGVSLSSQEAETNFRPREYLNINONS 1110

```

```

Qy      1178 VKQVFAFGAVENPEYLTPOGGAAPQHPPEPAPSFADNLYWDODPREGAPPESTFGK 1237
Db      1111 L---PLVSSGSMDDPDI---QAG-----YQAAF-----LPQALGALGNMF 1145
Qy      1238 TPTAENPEYLG 1248
Db      1146 LPAENLEYLIG 1156

RESULT 9
A:6223
kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; 159164
R:Kraus, M.H.; Iseling, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERB3, a third member of the ERBB/epider
A:Reference number: A36223; MUID:90083234; PMID:268785
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <RFA>
A:Cross-references: GB:M29366
R:Picman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor recept
A:Reference number: 159164; MUID:90311312; PMID:2164210
A:Accession: 159164
A:Status: preliminary
A:Molecule type: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:G183990; PID:AAA35979.1; PID:G306841
A:Gene: GDB:ERB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match      34.9%; Score 2380.5; DB 2; Length 1342;
Beet Local Similarity 40.0%; Pred. No. 7.1e-91;
Matches 526; Conservative 195; Mismatches 454; Indels 141; Gaps 35;

Qy      10 GLLALPFGAA--STQVCTGTDMKRLPLASPEHLDMLRLYQGCQVYQNLLEYLPT 67
Db      11 GLFSLARGSEVNSQAVCPGTLNGLSVTGAENQVQLYLYERCEVWGNLEIVLIGH 70
Qy      68 NASLSFLDIDIEVQGYVLIANQVRYQLRLRYRGQLFEDNYALAVLDNGPILNNTT 127
Db      71 NADSLFLQWIREVGYVLIANNEFSTPLPLRLRVARGVQVADGKFAIFVM-----LNNT 125
Qy      128 PVTGSPGGLRELOLRSLTEILKGVLIQRNPOLCYODTIMKDIHFKNOLATLIDTN 187
Db      126 ----NSSHALRQLHLTQVTEILSGVYIEKDKXCHMDTIMRIVADRO---AEIVKD 178
Qy      188 RSRACHPCSPMKSGRCWSESSEDCOSLTRTVCAAGC-ARCKGLPDPDCHEQCAAGTG 246
Db      179 NGRSCPPCHEVCKG-RCMGPSSEDCQTLTKTICAPQCGNHFGRPNQCCDBACAGGSG 237
Qy      247 PKHSDCLACLHFNNSGICELHCPALVYNTDTPFSMPREBRYTFGASCYVACPYNLYST 306
Db      238 PDLTDCFRCHFNDSGACVPRCPQLVYNTKTLFQLEPHNPKTYQYGVGCVASCNHYV-V 296
Qy      307 DVGSCCTVCPHNOEVTADGTQCEKSKPCARCVGLQMGVYIKANSKP--IGITEL- 363
Db      297 DQTSVCRACPDKXKLEVD-KNGLKWCCEPGGLCPACACGTG-----SGSRFQTVSSNIDG 350
Qy      364 FAGCKTIFGSLAPLPSFDGDPASNTAPLOPELOVFEETLEITGYLYISAMPDLSLDS 423

```

Db 351 EVNCTKILNLDPLITGLNGDPWPKI PALDPEKLVNFTVREITGYLNIQSPRHMHNF 410  
 QY 424 VFQNLQVIRGRIILHNGAVS-LTLOGLSIWMGLRSRLGSLALIHNTLHCFVHTVPM 482  
 Db 411 VTSNLTIGRSRLYNGFSLIMKNTLVNLSLGRSLKEISAGRIYISANRQLCYHNSLNM 470  
 QY 483 DQLFRRPHQALHTA-NRPEDEVGGLACHOLCARGHGCMWGPPOCVNCSQFLGQEGCV 541  
 Db 471 TKYLRGPTEBERLDIKNRRPRDCAVGRKVDPLCSSGGCGPQGLCSNYSRSGVCV 530  
 QY 542 EECRVIOGLPREYVNAHCLPCHPECOPONGSVTCGPADOCVACAHYKDPFVCARCP 601  
 Db 531 THCNFLNGERERFHAECSCSPRECQPMEGATATCGSSSDICACAHFRDHPHCVCSP 590  
 QY 602 SGVPRDLSTYPIKPPDEEGACOPPCINCHSCVDLDDKCGCAEQGA---SPLSYVA 657  
 Db 591 HGVLG--AKGPYKYPDVQNECRPHENCTOGCKPELDDCGQTLVLGKTHLTALTV 648  
 QY 658 VVGILLV-VVLGVVF---GILFNNFTVSFWLVRPKVASHLPELTPSGAMPNQAKRI 712  
 Db 649 IAGLVIFMVLGGSTFLYWRGRRIQNRARRYLE---RGESEIPLDPS-EKANKVLARI 703  
 QY 713 LKETLRKVKVIGSGAGFYVYKGIWIPDGENVKIPIVAKYLRNTPFKANKELDEAYVM 772  
 Db 704 FKETEIRKIKLVGSGVFGIVHKGVMIPEGESIKIPVCIKVIEDKSGQSGFOAVTDMAL 763  
 QY 773 AGVSPYVSRLLGICLTSTVQVLTOLMPYGLLDHRENRGRGSGODLNMGCQIKGMS 832  
 Db 764 GSDIHAIVRLGICPSSSIQTLVQYPLGSLLDHVRQHGALGPOLLNMGQIKAGVY 823  
 QY 833 YLEEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDDETEYHADGKVPKIMWLES 892  
 Db 824 YLEHGMVHRNLAAARNVLLKSPQVQVADGVADLLPPDDKQLLYSEAKPIKMWLES 883  
 QY 893 LRRRTFQSDVMSVGTVMELMTFGAKPYDGIAPAREPDLLEKGEPLPOPCITIDVMI 952  
 Db 884 HFGKYTHQSDVMSVGTVMELMTFGAPAGLAEPLDEKGERLAQOICTIDVMI 943  
 QY 953 MVKGMWIDSECRPFRELVSEFSRMAPDQRFVITQNEDELPA---SPIDSTFYRLLED 1009  
 Db 944 MVKGMWIDENIRPEKELANFTPMADPPRYLVIKES-GPGIAPRPHGLJNKKLE 1002  
 QY 1010 DDMGDLVDAEYLVPOQGFPCPPAPAGAGMVRHNSSTTRSGGGLTLGLEP-SEEA 1068  
 Db 1003 VELEPELDLDLEAEED---NLATTLTGSAISLPVGLTKMRPG 1043  
 QY 1069 PRSPLASSEGASGVFDGLGMGAKGLQSLPTHDPSPLORYSEDPVLP-----SE 1121  
 Db 1044 SQSLSPSSGV-MWNGNNGESQCESAVSGSSERCPVSLH-----EMPRGLASES 1096  
 QY 1122 TDGVYA-----PLTCSPOPE---YNQPDVROPSPREGP----- 1154  
 Db 1097 SEGHVTSSEAELOEKVMCSRSRSRPRPGSAYHSQSHSLTTPTPSPFGLSEEDV 1156  
 QY 1155 ---LPAAPAGATLEBAKTLSP-GKNGVY---KDVAFGGAVNPEYLPPOGGAAP 1203  
 Db 1157 NGYMPDTHLKGTSREGTSSREGTSSVGLSVLSTEDEDD-----EYEVNMRRRHSP 1208  
 QY 1204 QPAPPAFSPAFLNLYWD-----QDPERGAPPSSTKGIPTANPEYL 1247  
 Db 1209 -PHPPRSSLEELGYEYVDVGSLSLGSFOSCPHPPVIMPTAGTTPDEDEYVM 1263

## RESULT 10

JC4387  
 epidermal growth factor receptor homolog precursor - rat

N/Aternate names: ErbB3 protein; HER3 protein

C/Species: Rattus norvegicus (Norway rat)

C/Date: 17-Jan-1996 #sequence\_revision 19-Apr-1996 #text\_change 13-Nov-1998

C/Accession: J04387  
 R/Hellier, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.  
 Gene 165, 219-284, 1995  
 A/Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.

A/Reference number: J04387; MUID:96096535; PMID:8522190  
 A/Accession: J04387  
 A/Molecule type: mRNA  
 A/Residues: 1-1339 <HEL>  
 A/Cross-references: GB:U29339; NID:9915389; PID:9915390  
 A/Experimental source: liver  
 A/Note: The authors translated the codon AAC for residue 369 as Thr and GTT for res;  
 C/Comment: This protein is a functional heregulin receptor that transduces signals  
 A/Genetics:  
 C/Genes: ErbB3  
 C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h  
 C/Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prote  
 F.1-19/Domain: signal sequence #status predicted <SIG>  
 F.20-1339/Product: epidermal growth factor homolog #status predicted <MAT>  
 F.640-659/Domain: transmembrane #status predicted <TM>  
 F.705-970/Domain: protein kinase homology <KIN>  
 F.713-721/Region: protein kinase ATP-binding motif  
 F.939,1051,1156,1194,1196,1219,1257,1273,1286,1325/Binding site: phosphate (TY  
 Query Match 33.6%; Score 2292.5; DB 2; Length 1339;  
 Best Local Similarity 40.2%; Pred. No. 2.9e-87;  
 Matches 517; Conservative 171; Mismatches 434; Indels 165; Gaps 36;  
 QY 3 LAALCRWGLLALLPRGA---STVCTGTMTKRLRPSRPTHMDLRHLYOGGOVQGN 59  
 Db 7 LQVLC---FLSLARSGEMNSQAVCGTINGLSVGDADNQYQTLKLYEKEGVWGN 62  
 QY 60 LELTFLPTNASTFLFOIOEVQGYVLIHNOVROVPIQRLRIYVGTQLFEDNVALAVLDN 119  
 Db 63 LEIVLTGNADLSFLQWIREVTAYLVAMNEFSLPRLNLAHVAGTQYDGKPAIFV-- 120  
 QY 120 GDPINNTPTVGTASPGGLREIQLRSLEIFLKGVLICQNPOLCQDITLMDIETHKNNOL 179  
 Db 121 ---LNYNT---NSSHALRQLKFTQTLTILSGVYIEKNDKLCHMDTIDWDIYVR-- 170  
 QY 180 ALTLIDTNRSAHCPSMCKGRCMGWSSSDCOSLTPRYVAGGC-ARCKGRLPTDCHE 238  
 Db 171 GAEIVKNNAGNCPCHEVCKG-RCMGSGPDDQLITITICAPQCNCGCFEPNPGCHD 229  
 QY 239 QCAAGCTPKKSDCLACLHFNHSGICELHCPALVTYNTDTESMNPREGRTYFGASCYTA 298  
 Db 230 ECAGGCGSGHDDCFACRFRNDSGACVRCPEPLVYNNKLTQLEPNPHTKYOYGGVCAVS 289  
 QY 299 CPYNNLTDSGCTIVCLHNGEYVAEAGTQCEKSCPKCARVYGL-GMOYIYANKF 356  
 Db 290 CPNHFV-VQGFVCRACPPDKKEVD-KHGLMKCECCGGLCPKACGTSGSRYQVDSN 347  
 QY 357 IGITELFAGCKKIFGSLAFLPESFGDPASNTAPLOEOLQVETLEITGYLYISAP 416  
 Db 348 ID---GFVNCCTKIIGNDPLITGLNDPMMHKIPALDEKLVNFTVREITGYLNIQSWP 403  
 QY 417 DSLPDLISVQNLQVIRGRIILHNGAVS-LTLOGLSIWMGLRSRLGSLALIHNTLHC 475  
 Db 404 PMHNFVSFSLNLTIGRSLYNGFSLIMKNTLVNLSLGRSLKEISAGRIYISANQQLC 463  
 QY 476 FVHTVMDOLFRRPHQALHTA-NRPEDEVGGLACHOLCARGHGCMWGPPOCVNCSQF 534  
 Db 464 YHSLNMTVLLGSPBEERLDIKYRPLGCLAEKGVCOPLGSSGCGMGAPOGLCSNRY 523  
 QY 535 LRQDCVEECRYLQGLPREYVNAHCLPCHPECOPONGSVTCFGEADOCVACAHYKPP 594  
 Db 524 SREGVCVTHCNPLQEPREFVHAQCFSCHPRECLPMEGISTYNGSGSACACAHFRDP 583  
 QY 595 FCYARCPGKVPDLSYPIKPPDEEGACOPPCINCHSC--VDLDDKCGCAPEOASPLT 652  
 Db 584 HCVNSCPHILG-AKGPYKFPDAQNERPHENCTOGCNPRELDDCGQAEVLMSKH 641  
 QY 653 SIYSAVYGI-LLVVLGVVF---GILFNNFTVSFWLVRPKVASHLPELTPSGAMPNQ 707  
 Db 642 LVLAATVGLAVILMILIGSFLYWRGRRIQNRARRYLE---RGESEIPLDPS-EKANK 696  
 QY 708 AQMRILKETELRKVVLGSGAGFYVYKGIWIPDGENVKIPIVAKYLRNTPFKANKELD 767

Db 697 VLARIKETELRLKLVAGSGVFGTVAKG:WIPEGESIKIPVICIKVIEDKSGRQFQAVTD 756  
 QY 768 EAVVAVGVSPIYSSRLGICLTSTVOLVQLMPYGCGLDHPVENRGLSSQDLWMCQI 827  
 Db 757 HMLAVSLDHAHIVRLIGLCPGSSLOLVQYPLGSLDHYKHRETTLPOLLWVGQI 816  
 QY 828 AKMSVLEEDVRLVHRDLAARNVVKSPNHNKRTDFLARLLDIDEFYADAGKVIKMM 887  
 Db 817 AKMYLLEHSMVHRDLARNVVKSPQVQVADFEVADLLPDDQQLHSEAKPIKMM 876  
 QY 888 ALESILRRFTTHQSDVMSYGVTVWELMTFGAKPYDGI PARFIPDLLEKGERLPDPICIT 947  
 Db 877 ALESIFGKXTHQSDVMSYGVTVWELMTFGAEVYAGLRLEIPDLLEKGERLAQPOICIT 936  
 QY 948 DVMIMVKKCMIMIDSECRPRPRELVEFSFMAPDPQRFVYIOMEDIGPASPDLSTFYSRL 1007  
 Db 937 DVMIMVKKCMIMIDENIRPTFKELANFTKADPPRYLVIKRAS -GPGTF--PAAPPSVL 993  
 QY 1008 EDDMDGLDVAEEYLVPOQGFPCPDPAAGAGVHHRHSSSTRSGGDLTLGLEPSEE- 1066  
 Db 994 TKTEL-----QEALELEPEL-----DLDLIDLEAEEEG 1019  
 QY 1067 -----EAPRSPLAPSEG-----AGSDVPFGDLGMAAKGLQSL 1099  
 Db 1020 LATSLSAVALPFTGILTRPRGSQLSPSSGYWPMNQSSLGACLDLSAVLGGREQFSRPT 1079  
 QY 1100 PTHDSPPLQRYSEDPVPLPSETDGYV---APL-----TC-----SPOPE---YVN 1139  
 Db 1080 SLH-PIPRGR-----PASESEGHVYTGSEALGKVASVCRRSRSPRRRGDSAYHS 1131  
 QY 1140 QPVRPOPSPSPRGR-----LPAPRPAAGATLEAKTISP-GKRGVY-----KD 1181  
 Db 1132 QRHSLLTPTVPLSPGLEEDNGYVMPDTHLRGASSRREGTLSSVGLSGVLTGEDEED 1191  
 QY 1182 VFAFGAVENPEXYLTPQGAAPQHPHP 1208  
 Db 1192 -----EEVEYMKRKGSP-PRFP 1209  
 RESULT 11  
 TVFVLV  
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus  
 N:contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase  
 C:Species: avian leukosis virus, ALV  
 C:Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 11-Jun-1999  
 C:Accession: B00643; A00643  
 R:Nilsson, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.  
 Cell 41, 719-726, 1985  
 A:Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro  
 A:Reference number: A00643; MUID:85228222; PMID:2988784  
 A:Accession: B00643  
 A:Molecule type: mRNA  
 A:Residues: 1-698 <NID>  
 A:Cross-references: GB:M1006; GB:M1381; NID:9211749; PID:AAA48763.1; PID:9211750  
 A:Note: In Genbank entry CHKEBHF, release 109.0, the source is designated as Gallus gal  
 C:Comment: This protein is synthesized as a gag-env-erbB protein.  
 C:Genetics:  
 A:Gene: gag-env-erbB  
 A:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P  
 F:1-6/Products: gag protein (fragment) #status predicted <GAG>  
 F:7-59/Products: env protein (fragment) #status predicted <ENV>  
 F:60-698/Products: protein-tyrosine kinase erbB #status predicted <ERB>  
 F:194-459/Domain: protein kinase homology <KIN>  
 F:202-210/Region: protein kinase ATP-binding motif  
 F:229/Active site: Lys #status predicted

Query Match 24.8%; Score 1689.5; DB 1; Length 698;  
 Best Local Similarity 50.8%; Prid. No. 1e-62;  
 Matches 363; Conservative 78; Mismatches 149; Indels 125; Gaps 18;

QY 576 GPEADQCVCARHKDPFCVACRCPSGVKPDLSYMPKFPDEGACQPCPINCTHSQVDL 637

Db 60 GP--DHCKCAHFIDGPHCVACAPAGVLGENDTL-VWKYADANAVCOLCHPNCTRGCKGP 116  
 QY 638 DDKCPAQRASPLTSIVSANV-GILLVYVGVVFGILIFNNFV-SFWLRFVPKVASHL 695  
 Db 117 GLEGPP--NSKTPSIAAGVAGGLCLVAVGLDGLYLRHRIVRKTLRLLOERELV 173  
 QY 696 EPLTPSGAMPQAOQRILKETELRKVKVLSGAGTGVYKGIWIDGENVKIPVAKVLE 755  
 Db 174 EPLTPSGAMPQAOHRIKETEFKKVYLSGAGTGVYKGIWIDGENVKIPVAKVLE 233  
 QY 756 NTPSKANKELLIDEAVVAVGVSPIYSSRLGICLTSTVOLVQLMPYGCGLDHPVENRGL 815  
 Db 234 ATSPRANKELLIDEAVVAVSVNPHVCRLGICLTSTVOLVQLMPYGCGLDHYIREHKNI 293  
 QY 816 GSODLNMCMQIAGMSVLEEDVRLVHRDLAARNVVKSPNHNKRTDFLARLLDIDEFY 875  
 Db 294 GSQYILNMCMQIAGMYLERLRLVHRDLAARNVVKTPQVYKTTDFLARLLDIDEFY 353  
 QY 876 HADGKVIKMMALESILRRFTTHQSDVMSYGVTVWELMTFGAKPYDGI PARFIPDLLEK 935  
 Db 354 HAEGKVIKMMALESILHRIYTHQSDVMSYGVTVWELMTFGSKPYDGI PASEISSVLEK 413  
 QY 936 GERLPQRPICITDVMIMVKKCMIMIDSECRPRPRELVEFSFMAPDPQRFVYIOMEDIGP 994  
 Db 414 GERLPQRPICITDVMIMVKKCMIMIDADSRPFRELIAFSFMAPDPRYLVIOGDERHL 473  
 QY 995 ASPLDSTFYRSLLEDDMDGLDVAEEYLVPOQGFPCPDPAAGAGVHHRHSSSTRSGG 1054  
 Db 474 PSPDTSKRYRFLMEEDMEDIVDADEVYVPHQGF-----NSPST----- 513  
 QY 1055 GDULTGLEPSEEAARGL-----APSEGASDVFDGDLGMAAKGLQSLPTHDSPLQR 1109  
 Db 514 -----SRPLLSLSATSNKSNATNID-----RNGGHVVRDSFQR 551  
 QY 1110 YSEDPVPLPSET--DGVAAPLTCSPQPEYVQCDVRRPQSPREGPLPARPAGATLER 1167  
 Db 552 YSSDPTGNFLSESIDDGL-----PAEYVNO--LMPKKS----- 585  
 QY 1168 AKTSPGKGVKQVDF-----AFGAVENPEXYLTPQGAAPQHPHPAFS 1212  
 Db 586 --TAMVQNGIYNNISLTAISKLPMSRYQSHSTAVDNPEYL-----NTQSPFLAK 634  
 QY 1213 PAFNDLYWDC-----DPPE-----RGAPSPFTKQPTENPEYGLDVP 1252  
 Db 635 TVFESSPYIQSGNHQINLNDPDIQODFLPNETKPLGLKVPALBNPEYLVVAP 689  
 RESULT 12  
 TVYUH  
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain  
 C:Species: avian erythroblastosis virus  
 C:Date: 18-Apr-1984 #sequence\_revision 18-Apr-1984 #text\_change 11-Jun-1999  
 C:Accession: A00644; A38022  
 R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.  
 Cell 35, 71-78, 1983  
 A:Title: The erbB gene of avian erythroblastosis virus is a member of the src gene  
 A:Reference number: A00644; MUID:84026539; PMID:6313229  
 A:Accession: A00644  
 A:Molecule type: DNA  
 A:Residues: 1-604 <YAN>  
 A:Cross-references: GB:K01216; NID:G209676; PID:AAA42400.1; PID:G209678  
 R:Deblaire, B.; Henry, C.; Benatissa, M.; Bisette, G.; Clavertie, J.M.; Saulle, S.; Mar  
 Science 224, 1456-1459, 1984  
 A:Title: Sequencing the erbB gene of avian erythroblastosis virus reveals a new typ  
 A:Reference number: A38022; MUID:8422957; PMID:6328658  
 A:Accession: A38022  
 A:Molecule type: DNA  
 A:Residues: 1-28, 'W', '30-139', 'V', '141-145', 'V', '147-152 <DEB>  
 A:Cross-references: GB:K02006  
 C:Genetics:  
 A:Gene: erbB  
 A:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specifi



F:130-395/Domain: protein kinase homology <Kin>  
F:138-146/Region: protein kinase ATP-binding motif  
F:165/Active site: Lys #status predicted

Query Match 23.9%; Score 1626; DB 1; Length 604;  
Best Local Similarity 50.7%; Pred. No. 3, 5e-60;  
Matches 349; Conservative 74; Mismatches 140; Indels 126; Gaps 16;

QY 587 CAHYKDPFCVAPCPGKVPDLSYMPKRPDEGACQPPINCTHSQVLDKCGCPAQ 646  
DB 3 CAHFIDGPHCVKACPAVGLGENDTLVRKYADANAACOLCHNCTRGCKGPGEGCP 58  
QY 647 BASPLTSIVAVV-GILLVVVIGVVGIIIFNNFTV-SPMLRPKVASASHLEPLTPSGAM 704  
DB 59 NSKTPSIAGVGLGLVVGIGLVIRRHRYKTKTLRLLOERELVEPLTSSGA 118  
QY 705 PNOQMRILKETELRKVKVLSGAFGVYKGIWIPDGENYKIPVAIKVIRENTSPRANKE 764  
DB 119 PNOAHLRIKETEFKKVKVLSGAFGVYKGIWIPDGENYKIPVAIKVIRENTSPRANKE 178  
QY 765 ILDEAYVMAGVSPYVSRLLGLTSTVQVLTQMLMPYGLDHRNRRGLSGODLLNMC 824  
DB 179 ILDEAYVMASVNDPHCRLLGLTSTVQVLTQMLMPYGLDHRNRRGLSGODLLNMC 238  
QY 825 MOIAKMSYDEDELVHRDLAARNVLKSPNHVKITDPGLARLLDIDETEHADGGKVP 884  
DB 239 VOIAKGMNVLERRLVHRDLAARNVLKTPQHVKITDPGLAKLLGADKEXYHAGKVP 298  
QY 885 KMALESILRRRTHOSDWSYGVYVWELTFGAKPPDGI-PAREIDLEKGRLLPOPI 944  
DB 299 KMALESILRRRTHOSDWSYGVYVWELTFGAKPPDGI-PASEISSVLEKGRLLPOPI 358  
QY 945 CTIDVYIMYKCMWIDSECRPRELVEFSRMAPDPORFVVIQ-NEDIGPASPLDSTFY 1003  
DB 359 CTIDVYIMYKCMWIDSECRPRELVEFSRMAPDPORFVVIQ-NEDIGPASPLDSTFY 418  
QY 1004 RSLIEDDMDGLVDAEYVLPQGFPCPPAPAGAGVHHRRSSSTRSGDULTLGLEP 1063  
DB 419 RSLIEDDMDGLVDAEYVLPQGFPCPPAPAGAGVHHRRSSSTRSGDULTLGLEP 449  
QY 1064 SEEARPSPL-----ASEGASGDPFDGLGMAKAGLQSLPHDPEPLQRYSEDPVPL 1118  
DB 450 -----SRPLSLSATSNNSATKCTD-----RNGQHPREDSFVRYSDPTGNE 496  
QY 1119 PSET-DGYVAPLTCSQPEYVNPQVPPSPFREGPLPAPBPAGATLERAKTLPKGN 1176  
DB 497 LEESIDDGFL-----PAEYVNPQ-LMPKKPSTAM----- 524  
QY 1177 GVYKDVFAF-----GGAVENPEYLTQGGAAQPPHPAPAPAFDN 1217  
DB 525 -VONQIYNFISLTAKLPMDSRKONSHSTAVDPEYL-----NTNOSPLAKTVFES 575  
QY 1218 LYYWDQDPPERGAPSTFKGTPTAENPEY 1246  
DB 576 SPYWIQSGNHQ-----INDNDY 594

RESULT 13  
QYFE  
epidermal growth factor receptor - fruit fly (Drosophila melanogaster)  
N/Contains: protein-tyrosine kinase (BC 2.7.1.112) erDB  
C/Species: Drosophila melanogaster  
C/Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 11-Jun-1999  
C/Accession: A00640; A38021  
R/LYNEH, E.; GLAZER, L.; SEGAL, D.; SCHLESINGER, J.; SHILO, B. Z.  
Cell 40, 559-607, 1985  
A/Title: The Drosophila EGF receptor gene homolog: conservation of both hormone binding  
A/Reference number: A00640; MUID:85124611; PMID:2982499  
A/Accession: A00640  
A/Molecule type: DNA  
A/Residues: 1-1330 <Lit>  
A/Cross-references: EMBL:K03054  
R/Madsen, S. C.; Vincent III, W. S.; Bilodeau-Wentworth, D.

Nature 314, 178-180, 1995

A/Title: A Drosophila genomic sequence with homology to human epidermal growth factor  
A/Reference number: A38021; MUID:85137938; PMID:2983232

A/Accession: A38021

A/Molecule type: DNA

A/Residues: 'A', 832-866, 'V', 868-943, 'QTPSLVK' <MAD>

A/Cross-references: EMBL:X02293; NID:97922; PIDN:CAA26157.1; PID:9929565

C/Comment: This sequence is tentative because the introns have not been identified.

C/Genetics:

A/Genes: Flybase:Bgfr

A/Cross-references: Flybase:FBgn003731

A/Map position: 2 57P

C/Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; ph

F:1-732/Domain: extracellular #status predicted <EXT>

F:733-764/Domain: transmembrane #status predicted <TM>

F:765-1330/Domain: intracellular #status predicted <INT>

F:808-1072/Domain: protein kinase homology <Kin>

F:816-824/Region: protein kinase ATP-binding motif

F:122-300,324,363,518,688,695,700/Binding site: carbohydrate (Asn) (covalent) #stat

F:172/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predic

F:183/Active site: Lys #status predicted

F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status pr

Query Match 23.5%; Score 1599.5; DB 1; Length 1330;  
Best Local Similarity 29.3%; Pred. No. 9, 4e-59;

Matches 406; Conservative 177; Mismatches 415; Indels 389; Gaps 40;

QY 80 VQGYVLAMNQVRQVRLRIIVGTQF-----EDNALVALVDGPELNTTPTVGTASP 134  
DB 38 ITNYVIGLDLPCTLSRLQIRKRTLSVSEBKALTY-----TY 81  
QY 135 GGLREQLRSLTEILKGVLLIQNPOLCYDTIIMKDI-FHKNQALATLLIDNRSRACHP 194  
DB 82 SMYTLLEIPDLRDVLNQGVSFNNYNLCHEMTIQSBIVNSGTADYVYVDFAPARECP 141  
QY 195 CSPMKGSRCKWSESEDDQSLTRYCAGCA--RKGLPLPDCHCECAAGTGRHSDC 252  
DB 142 CHESQTHG-CWGEKGNQKFSKLTCSQCGRGYGRKPCCHLFAAGGTGTQKDC 200  
QY 253 IACLPHNSGICELCPALVTYNTDFESMPREGRYFGASCYVACEVNYLSTVGSCT 312  
DB 201 IACKNPFDEAVKECPKPKKYNPTTYLTENPEKVAVGATCVACEP-GHLIRNGACV 259  
QY 313 LVCPLHNOETVAEDGTQRCCKSPCAVYCGLMQYIKANSKFGITEL-----EPAG 366  
DB 260 RSCPDQKMDKGE-----CVPNGCPCKPTC-----GVTVLHAGNIDSEFN 300  
QY 367 CKKFGSLAFRESFDG--DPASNTA-----PLQEPOLQVETLEITGYIISAMPDS 418  
DB 301 CTVIDGNITILDQTSRGPDVYANTMGPRYIFLDPERREVSIVKEITGYINIGTHPQ 360  
QY 419 LPDLVFNQLQVIRGRILHNGAY-SLTLOGIGISWLGRLSELGSLALHHNTHLCFV 477  
DB 361 FNLSTFRLLETIRHQLMESMFALATVKSLSVLEMRNLQISSGSVVIQHNRLDLCV 420  
QY 478 HTVPMDQLFRNPHQALHTANRPEDEC----- 504  
DB 421 SNIRWPAIQKEPEQKWNENLRADLCFKFTLLISVGNITIMHFAICREKMHLLSV 480  
QY 505 ----- 504  
DB 481 QGRLLGSHGSAVPLYQLQFQHLHRLMYIQVINSIQKSNHQLTDACSPSVPT 540  
QY 505 -----VG 506  
DB 541 SLTIERAYAIOSAGLAMELEQITARSASMRHSKTLPAEGROVPRWVFLGVCASAPAGIA 600  
QY 507 EGLA-----CHQCARGHGCPGPTQCVNCSQFLRGGECVEGRVQLGPREVY--N 556  
DB 601 EPLAGAVARKCHPLCELCTNYGYHGOVSKCTHYKRRQGTCC-----PADYTDSE 654  
QY 557 ARHCLPCHPECCQPNQSVTCFPEPADQVACAHYK-----DPFF-----CVARCSG 603



```

Db      655 QRECFGRHPEC---NG---CTGGADCKSCRFKLPANETGPRVNSMTMFCSTKCPLE 708
Qy      604 VR-PDLSYPMFKPFBEGACQPCPINCCHSCVDLDCQPAQORASPTLSVSAVGL 662
Db      709 MRHVNQYTAIGPY-----CAASPRSSKITANDL-----VNNFIITIGAV 749
Qy      663 LVVVLGVFGILF-----NNFTVSPMLRVKPVASHLEPTLSPGAMPNOAKRILE 715
Db      750 LVPTICLCVVTYICQKQAKKAYKMMALSGRDS--EPLRPENIGANLCKLRIVVD 807
Qy      716 TELRKVKVLSGAFGTYYKGIWIPDGENYKIPAIKVLRENTSPKANKEILDEAYVAGV 775
Db      808 AELRKGVLMGMAFGYVGVKGVWVEGENYKIPAIKELLKSTGABSESEFLREAYIMASE 867
Qy      776 GSPYVRILGICLSTVQVLTQMLPFGCLLDYKNNRDKIGSKALLNMSTQJAKMSYLE 835
Db      868 EHVNLKLAVCHSSQMLTQMLPFGCLLDYKNNRDKIGSKALLNMSTQJAKMSYLE 927
Qy      836 DVLVHRLAARVLYK---SPNHVKITDGLARLLDIDETEXHADGKVPIMKMALES1 892
Db      928 EKLVHRLAARVLYK---SPNHVKITDGLARLLDIDETEXHADGKVPIMKMALES1 983
Qy      893 LRRRFTHOSDWSYGVYWEIMTFGAKPYDGIPIAREIPDLLEKGEPLPQPICTIDVMI 952
Db      984 RNVFTSKSDVMAFGVYIMELLTFGQRPHEHNPAKDIPDLIEVGLTEQPEICSLDIYCT 1043
Qy      953 MYVCMWIDSECRPFRELYSEFMRARDPQRFVIONEDLG--PASPLDSTFYRSLIEDD 1010
Db      1044 LLSCHMLDAMRRTFQGLTVFAEFARDGRIYALIGDKFTRIPA-----YTSQDEMD 1096
Qy      1011 ---DMGDLVDAEYLVPOQGFPCPDPAAGKGVHHRHSSSTRSGGDLTGLPSEEE 1067
Db      1097 LIRKLAPTDGSFAIKKPDYLOPKALGPS-----HRTDCI-----DE 1135
Qy      1068 AP-----RSPLASGAGSDVDPG---DLGMAKAGLSLPTHDSPLOKRYSEDPVPL 1118
Db      1136 MFLANRYCQDPKSKHNSGTGDDERDSAREVGVNLR-----LTL 1174
Qy      1119 PSETDGYVAFLTCSPOPEYVNOQDVRPQPSPREGLPAPAPAGATLEAKTILSPKNGV 1178
Db      1175 PVDEDDYLMF--TCQPPNNNNNNNN-----NPNQNNMAVGVAGYV----- 1214
Qy      1179 VKGVFAFGAVENREYL---TPQGAAPQPH-----PPRASP-AF 1215
Db      1215 --DLIGVPSVDNPEYLNAQTIGVSESPITQTGTGIPVNGGCTMEVAVPMPGSEPTSS 1272
Qy      1216 DNLVYWD 1222
Db      1273 DHEVYND 1279

```

## RESULT 14

S35745  
 protein-tyrosine kinase (EC 2.7.1.112) erbb - avian erythroblastosis virus  
 C/Species: avian erythroblastosis virus  
 C/Date: 03-Mar-1994 #sequence\_revision 26-May-1995 #text\_change 28-Feb-1997  
 C/Accession: S35745  
 R/Vennstrom, B.  
 submitted to the EMBL Data Library, March 1993  
 A/Reference number: S35743  
 A/Accession: S35745  
 A/Molecule type: DNA  
 A/Residues: 1-544 <VEN>  
 A/Cross-references: EMBL:X12707  
 C/Genetics:  
 A/Gene: erbb  
 C/Superfamily: epidermal growth factor receptor; protein kinase homology  
 C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P  
 F/135-400/Domain: protein kinase homology <KIN>  
 F/143-151/Region: protein kinase ATP-binding motif  
 F/170/Active site: Lys #status predicted

```

Query Match      23.0%, Score 1570; DB 2; Length 544;
Best Local Similarity 53.3%, Pred. No. 6,46-58;
Matches 334; Conservative 68; Mismatches 133; Indels 92; Gaps 15;

Qy      578 GPEADQVACHAHYKDPPECVVARCPGKVDLSYPMFKPFBEGACQPCPINCCHSCVDL 637
Db      1 GP--DHCKKCAHFDIGPCHVACVAGVLENDTL--VWKYADANAVCQCHNCTRGCKGP 57
Qy      638 DQKGPARGASPLTSIVSAV--GILLVYGVVFGILIFNNFTV--SPMLRVKVASHL 695
Db      58 GLEGC---NSKPSIAGVYVGLLCLVVGILGILYLRHRHRYKRTLRLLDRELV 114
Qy      636 EPLTSGAMPNOAKRIKETELRKVKVLSGAFGTYYKGIWIPDGENYKIPAIKVLRE 755
Db      115 EPLTSGAMPNOAHLRIKETEFKVKVLSGAFGTYYKGIWIPDGENYKIPAIKVLRE 174
Qy      756 NTSPPANKEILDEAYVMAVGSVYRSLGICLSTVQVLTQMLPFGCLLDYKNNRRL 815
Db      175 ATSPKANKEILDEAYVMAVSDNPHVCRLLGICLSTVQVLTQMLPFGCLLDYKNNR 234
Qy      816 GSQDILNMQOIAKMSYLEDVRLVHRDLAARVLYKSPNHVKITDGLARLLDIDETEX 875
Db      235 GSQVILNMQOIAKMSYLEDVRLVHRDLAARVLYKSPNHVKITDGLARLLDIDETEX 294
Qy      876 HADGKVPIMKMALESILRRRFTHOSDWSYGVYWEIMTFGAKPYDGIPIAREIPDLLEK 935
Db      295 HADGKVPIMKMALESILRRRFTHOSDWSYGVYWEIMTFGAKPYDGIPIAREIPDLLEK 354
Qy      936 GERLPQPICTIDVYMIWVKCMWIDSECRPFRELYSEFMRARDPQRFVIONEDLG-- 994
Db      355 GERLPQPICTIDVYMIWVKCMWIDSECRPFRELYSEFMRARDPQRFVIONEDLG-- 414
Qy      995 ASPLDSTFYRSLIEDDMDGDLVDAEYLVPOQGFPCPDPAAGKGVHHRHSSSTRSGG 1054
Db      415 PSPDSTFYRSLIEDDMDGDLVDAEYLVPOQGFPCPDPAAGKGVHHRHSSSTRSGG 454
Qy      1055 GDLTGLPSEEEARSP-----APSGASDVDPDGLGMAKAGLSLPTHDSPLOKRY 1109
Db      455 GDLTGLPSEEEARSP-----APSGASDVDPDGLGMAKAGLSLPTHDSPLOKRY 481
Qy      1110 YSEDPVLPSETDGYVAFLTCSPOPEYVNOQDVRPQPSPREGLPAPAPAGAT-LEBA 1168
Db      482 -----PYREDFL-----PAPEVNG--LMPKPSRPMVNGQYNYISLAISKL 524
Qy      1169 KTLSPKNGVAKDVFAFGAVENREYL 1195
Db      525 PIDSRYN-----SHSTAVDNPEYL 544

```

## RESULT 15

S00727  
 kinase-related transforming protein (erbB) (EC 2.7.1.1.-) - avian erythroblastosis vi  
 C/Species: avian erythroblastosis virus  
 C/Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 28-Feb-1997  
 C/Accession: S00727  
 R/Scottling, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.  
 Oncogene Res. 1, 265-278, 1987  
 A/Title: Common site of mutation in the erbb gene of avian erythroblastosis virus m  
 A/Reference number: S00727; MUID:88217326; PMID:2897102  
 A/Accession: S00727  
 A/Molecule type: DNA  
 A/Residues: 1-545 <SCO>  
 A/Cross-references: EMBL:X06943  
 C/Genetics:  
 A/Gene: erbb  
 C/Superfamily: epidermal growth factor receptor; protein kinase homology  
 C/Keywords: ATP; phosphotransferase  
 F/135-400/Domain: protein kinase homology <KIN>  
 F/143-151/Region: protein kinase ATP-binding motif

Query Match 22.9%, Score 1563; DB 2; Length 545;  
 Best Local Similarity 53.3%, Pred. No. 1,36-57;  
 Matches 334; Conservative 67; Mismatches 134; Indels 92; Gaps 15;

```
QY 578 GPEADQCVACAHYKDPFCVACPSGVKDDLSIMPIWKFPDEBACQCPPICTHSCVDL 637
Db 1 GP--DHMKCAHFIDGPHCVKACPAVLGENDTL-VWKYADANAACOLCHPNCOTRCKGP 57
QY 638 DDGCEAFBQASPLSTISVAVV-GILLVVLGVVFGILLFNNFTV-SFWLRYVKVASHL 695
Db 58 GLEGCP--NGSKTPTSIAGVVGGLCLVVGIGIGLYLRRIHVKRTLRRLQRELY 114
QY 696 EPLTPSGAMPNQAOMRIKTELKRVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVIRE 755
Db 115 EPLTPSGEAPNQAHLRIKTEEFKVKVLFGAFTVYKGLWIPGEKVTIPVAIKELRE 174
QY 756 NTSPKANKETLDEAVYVAGVSPYVSRLLGICLTSTVOLVTOLMPYGCILDHRENRGL 815
Db 175 ATSPKANKETLDEAVYVAVSDNPHVCRLGICLTSTVOLITOLMPYGCILDYIREKDN 234
QY 816 GSODILNMCMQIAKMSYLEDEVLVHRDLAARNVLVKSPPNHVKITDGLARLLIDETEV 875
Db 235 GSQYLNMCMQIAKMMYLEERHLVHRDLAARNVLVKTPODVKITDGLAKOLGADEKEY 294
QY 876 HADGKVPITKMALESLIRRRFTHOSVWSYGVTWELMTFGAKPYDGI PAREIPDLLEK 935
Db 295 HAEGKVPITKMALESLIRRIYTHOSVWSYGVTWELMTFGSKPYDGI PASEISSVLEK 354
QY 936 GERLPQPICTIDVYIMVCKMMIDSECRPRELVSFSGMARDPQRFVVIQ-NEDLGP 994
Db 355 GERLPQPICTIDVYIMVCKMMSDASRPFRELIABFSKMARDPERYLVIGDERMHL 414
QY 995 ASPLDSTFYRSLLEDMDGLVDAEYLVPOQGFPCDPAPGAGMTHHRSSSTRSGG 1054
Db 415 PSPDTSKFRKRLMEEDMEDIVDAEYLVPHQGF-----NSPST---- 454
QY 1055 GDLTLGLPSEEEAPRSPL-----APSEGASDVFDGLGMGAAGLQSLPTHDPSPLQR 1109
Db 455 -----SRTPLLSLSLATSNSATINCIDRNG-----H----- 481
QY 1110 YSEDPVPLPSETDGYVAPLTCSPQPEYVNOVDVRFQPSFREGPLFAAPAGAT-LERA 1168
Db 482 -----FVREDGFL-----PAPRYVNO--LMPKKPSTAMVONQIYVYISLTAISK 524
QY 1169 KTISPKNQGVKDVAFAGAVENPEYL 1195
Db 525 PMDSRYQN-----SHSTAYDNPEYL 544
```

Search completed: July 22, 2003, 09:27:42  
Job time : 30.25 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.625 Seconds

(without alignments)  
4891.279 Million cell updates/sec

Title: SEQ4-653-675-14

Sequence: 1 MEALALCRMGLLALLPACA.....TRKGTPTAENPEYLGLDVPV 1253

Scoring table: BLOSUM62

Gapped 10.0 ; Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6604	96.9	1255	1	ERB2_HUMAN
2	5817	85.4	1257	1	ERB2_RAT
3	5801.5	85.1	1254	1	ERB2_MESAU
4	3067	45.0	1210	1	EGFR_HUMAN
5	3041	44.6	1210	1	EGFR_MOUSE
6	2908.5	42.7	1308	1	ERB4_HUMAN
7	2889	42.4	1308	1	ERB4_RAT
8	2614.5	38.4	1157	1	XRK_XIPMA
9	2389.5	35.1	1132	1	ERB3_HUMAN
10	2315.5	34.0	1132	1	ERB3_RAT
11	1924	28.2	1426	1	EGFR_DROME
12	1672.5	24.5	634	1	ERBB_ALV
13	1626	23.9	604	1	ERBB_AVIER
14	1553	22.8	540	1	ERBB_AVIEU
15	1518	22.3	703	1	EGFR_CHICK
16	1287	18.9	1323	1	LT22_CABEL
17	1142.5	16.8	245	1	ERB2_MOUSE
18	724	10.6	1363	1	ILPR_BRALA
19	706	10.4	1363	1	INSR_RAT
20	705.5	10.4	1372	1	INSR_MOUSE
21	705	10.3	1300	1	IRR_MOUSE
22	703	10.3	1382	1	INSR_HUMAN
23	692.5	10.2	1300	1	IRR_MOUSE
24	685	10.1	1257	1	IRR_HUMAN
25	685	10.1	1477	1	HTK7_HYDAT
26	680	10.0	1607	1	MITP_LYMAT
27	638	9.4	1367	1	IGLR_HUMAN
28	625	9.2	1373	1	IGLR_MOUSE
29	625	9.2	2146	1	INSR_DROME
30	622	9.1	1390	1	INSR_MOUSE
31	621.5	9.1	1370	1	IGLR_RAT
32	604	8.9	1114	1	RET_HUMAN
33	592	8.7	987	1	EPB4_HUMAN

34	584.5	8.6	984	1	EPB1_CHICK	Q07494	gallus gall
35	576.5	8.5	984	1	EPB1_RAT	P09759	rattus norv
36	573	8.4	902	1	EPB2_XENLA	O91736	xenopus lae
37	572.5	8.4	977	1	EPB2_MOUSE	O03145	mus muscu
38	572	8.4	987	1	EPB4_MOUSE	P54761	mus muscu
39	570.5	8.4	984	1	EPB1_HUMAN	P84762	homo sapien
40	569	8.4	757	1	HT16_HYDAT	P53356	hydra atten
41	569	8.4	1068	1	FAK1_XENLA	O00944	gallus gall
42	568.5	8.3	1053	1	FAK1_CHICK	P34152	mus muscu
43	567	8.3	1052	1	FAK1_MOUSE	O91571	xenopus lae
44	565.5	8.3	985	1	EPB4_XENLA	O35346	rattus norv
45	562	8.2	1055	1	FAK1_RAT		

## ALIGNMENTS

RESULT 1  
ID ERB2\_HUMAN STANDARD; PRT; 1255 AA.  
AC P04626;  
DT 13-AUG-1987 (Rel. 05, Created)  
DT 13-AUG-1987 (Rel. 05, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Receptor protein-tyrosine kinase erbB-2 precursor (BC 2.7.1.112)  
DE (p155erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell  
DE surface receptor (HER2) (MLN 19).  
GN ERBB2 OR HER2 OR NGL OR NEU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
OX NCBI\_Taxid:9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86118663; PubMed=3003577;  
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,  
RA Saito T., Toyoshima K.;  
RT "Similarity of protein encoded by the human c-erb-B-2 gene to  
RT epidermal growth factor receptor.";  
RL Nature 319:230-234 (1986).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=86070181; PubMed=2999974;  
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,  
RA McGrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,  
RA Francke U., Levinson A., Ullrich A.;  
RT "Tyrosine kinase receptor with extensive homology to EGF receptor  
RT shares chromosomal location with neu oncogene.";  
RL Science 230:1132-1139 (1985).  
RN [3]  
RP SEQUENCE OF 737-1031 FROM N.A.  
RX MEDLINE=86016729; PubMed=2995967;  
RA Semba K., Kanata N., Toyoshima K., Yamamoto T.;  
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the  
RT human salivary gland adenocarcinoma.";  
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501 (1985).  
RN [4]  
RP VARIANTS VAL-654 AND VAL-655.  
RX MEDLINE=93194196; PubMed=8095489;  
RA Ehsani A., Low U., Wallace R.B., Wu A.M.;  
RT "Characterization of a new allele of the human ERBB2 gene by allele-  
RT specific competition hybridization.";  
RL Genomics 15:426-429 (1993).  
CC -!- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.  
CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A  
CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
CC ALPHA AND AMPHIREGULIN.  
CC -!- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -!- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
CC (POTENTIAL).  
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY).  
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN  
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M11767; AAA35808.1; JOINED.  
 DR EMBL: M11761; AAA35808.1; JOINED.  
 DR EMBL: M11762; AAA35808.1; JOINED.  
 DR EMBL: M11763; AAA35808.1; JOINED.  
 DR EMBL: M11764; AAA35808.1; JOINED.  
 DR EMBL: M11765; AAA35808.1; JOINED.  
 DR EMBL: M11766; AAA35808.1; JOINED.  
 DR EMBL: M11730; AAA5493.1; JOINED.  
 DR EMBL: M12036; AAA35978.1; JOINED.  
 DR EMBL: X03363; CAA27060.1; JOINED.  
 DR PIR: A25491; A25491.  
 DR PIR: A24571; A24571.  
 DR HSSP: P11362; IFBK.  
 DR Genew: HGNC:3430; ERBB2.  
 DR MIM: 164870; ERBB2.  
 DR Interpro: IPR000494; EGFR\_L\_domain.  
 DR Interpro: IPR000719; Euk\_Dkinase.  
 DR Interpro: IPR002174; Furin-like.  
 DR Interpro: IPR001245; Tyr\_Dkinase.  
 DR Interpro: IPR004019; YIP\_motif.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR Pfam: PF02757; YIP; 2.  
 DR Prodom: PD000001; Euk\_Dkinase; 1.  
 DR SMART: SM00261; FU; 3.  
 DR SMART: SM00219; TyKc; 1.  
 DR Prosite: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR Prosite: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR Prosite: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Polymorphism.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 653 675 POTENTIAL.  
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 720 987 PROTEIN KINASE.  
 FT BINDING 726 734 ATP (BY SIMILARITY).  
 FT BINDING 753 753 ATP (BY SIMILARITY).  
 FT ACT\_SITE 845 845 BY SIMILARITY.  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 220 227 BY SIMILARITY.  
 FT DISULFID 224 235 BY SIMILARITY.  
 FT DISULFID 236 244 BY SIMILARITY.  
 FT DISULFID 240 252 BY SIMILARITY.  
 FT DISULFID 255 264 BY SIMILARITY.  
 FT DISULFID 268 295 BY SIMILARITY.  
 FT DISULFID 299 311 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 511 520 BY SIMILARITY.  
 FT DISULFID 515 528 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 587 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAc...) (POTENTIAL).  
 FT CARBOHYD 124 124 N-LINKED (GLCNAc...) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAc...) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAc...) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAc...) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAc...) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAc...) (POTENTIAL).  
 FT VARIANT 654 654 I->V.  
 FT VARIANT 655 655 /FTID=VAR\_004077.  
 FT CONFLICT 1170 1170 /FTID=VAR\_004078.  
 FT SEQUENCE 1255 AA; 137909 MW; 3989DFDA04DCF962 CRC64;  
 SQ  
 Query Match 96.9%; Score 6604; DB 1; Length 1255;  
 Best Local Similarity 96.7%; Pred. No. 0;  
 Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;  
 QY 1 METALCRWGLLALPPGAASVCTGTDMLRPLASPEHLDMLRLHLYGSCQVVGNL 60  
 DB 1 MELALCRWGLLALPPGAASVCTGTDMLRPLASPEHLDMLRLHLYGSCQVVGNL 60  
 QY 61 ELTYLPTNASLFLDIDIEVQGVILIANQVRQVPLQRLIRYRGQLFEDNVALVDNG 120  
 DB 61 ELTYLPTNASLFLDIDIEVQGVILIANQVRQVPLQRLIRYRGQLFEDNVALVDNG 120  
 QY 121 DELNNTPTPTGASPGGLRELQRLSTLEIKGVLQORNPOLCYOPTILMKDIFHKNQLA 180  
 DB 121 DELNNTPTPTGASPGGLRELQRLSTLEIKGVLQORNPOLCYOPTILMKDIFHKNQLA 180  
 QY 121 DELNNTPTPTGASPGGLRELQRLSTLEIKGVLQORNPOLCYOPTILMKDIFHKNQLA 180  
 DB 121 DELNNTPTPTGASPGGLRELQRLSTLEIKGVLQORNPOLCYOPTILMKDIFHKNQLA 180  
 QY 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCQSLTRIVYACGACRCAGPLPTDCHEQC 240  
 DB 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCQSLTRIVYACGACRCAGPLPTDCHEQC 240  
 QY 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCQSLTRIVYACGACRCAGPLPTDCHEQC 240  
 DB 181 LTLIDNRSRACHPCSPMKGSRGWESSEDCQSLTRIVYACGACRCAGPLPTDCHEQC 240  
 QY 241 AAGCTGPKKSDCLAFHNSGICELHCPALVTYNTDPESPNEBGRYTFASCVTACP 300  
 DB 241 AAGCTGPKKSDCLAFHNSGICELHCPALVTYNTDPESPNEBGRYTFASCVTACP 300  
 QY 301 YVYLSIDVGSCTLYVGLHNEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360  
 DB 301 YVYLSIDVGSCTLYVGLHNEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360  
 QY 301 YVYLSIDVGSCTLYVGLHNEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360  
 DB 301 YVYLSIDVGSCTLYVGLHNEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGIT 360  
 QY 361 ELEFACCKIFGSLAFPSFDDGPNASNPAPLOPOLVFEETLGTLYISAMPDSLIP 420  
 DB 361 ELEFACCKIFGSLAFPSFDDGPNASNPAPLOPOLVFEETLGTLYISAMPDSLIP 420  
 QY 361 IOEFACCKIFGSLAFPSFDDGPNASNPAPLOPOLVFEETLGTLYISAMPDSLIP 420  
 DB 361 IOEFACCKIFGSLAFPSFDDGPNASNPAPLOPOLVFEETLGTLYISAMPDSLIP 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGSRSLRELSGALIHNNHLCFVTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGSRSLRELSGALIHNNHLCFVTV 480  
 QY 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGSRSLRELSGALIHNNHLCFVTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYSLTQGLGISWLGSRSLRELSGALIHNNHLCFVTV 480  
 QY 481 PMDQLFRNPHQALLHNPANRPEDECVGEGACGQLCARGCMGPGPTQCNCGFLRGDC 540  
 DB 481 PMDQLFRNPHQALLHNPANRPEDECVGEGACGQLCARGCMGPGPTQCNCGFLRGDC 540  
 QY 481 PMDQLFRNPHQALLHNPANRPEDECVGEGACGQLCARGCMGPGPTQCNCGFLRGDC 540  
 DB 481 PMDQLFRNPHQALLHNPANRPEDECVGEGACGQLCARGCMGPGPTQCNCGFLRGDC 540  
 QY 541 VEECRVLQGLPREYVVARHCLPCHPECPQNSGVTGFEADQCVACAHYKPPCVAC 600  
 DB 541 VEECRVLQGLPREYVVARHCLPCHPECPQNSGVTGFEADQCVACAHYKPPCVAC 600  
 QY 541 VEECRVLQGLPREYVVARHCLPCHPECPQNSGVTGFEADQCVACAHYKPPCVAC 600  
 DB 541 VEECRVLQGLPREYVVARHCLPCHPECPQNSGVTGFEADQCVACAHYKPPCVAC 600  
 QY 601 PSGVRDLSTYMPWKPEDEGAQCPCPINCTHSVDLDKGCQPAERASPLSIISAVVG 660  
 DB 601 PSGVRDLSTYMPWKPEDEGAQCPCPINCTHSVDLDKGCQPAERASPLSIISAVVG 660  
 QY 601 PSGVRDLSTYMPWKPEDEGAQCPCPINCTHSVDLDKGCQPAERASPLSIISAVVG 660  
 DB 601 PSGVRDLSTYMPWKPEDEGAQCPCPINCTHSVDLDKGCQPAERASPLSIISAVVG 660  
 QY 661 ILVVVLLGVVFGILIIRROQKIRKYMRLLOTEL---VEPLIPSGMPPQAQMRILK 716  
 DB 661 ILVVVLLGVVFGILIIRROQKIRKYMRLLOTEL---VEPLIPSGMPPQAQMRILK 716

Oy	715	ETEJRKXVLGSAFCTYVKGIPIGGEVVKIPVAIKYLRNTSPKANKELIDAYVMAG	774
Db	717	ETELRKXVLGSAFCTYVKGIPIGGEVVKIPVAIKYLRNTSPKANKELIDAYVMAG	776
Oy	775	VGSBYVSRLIGLICITSTVOLVQLMPYGCILDHVENRGLGSDLLNMCQIAKGNSTYL	834
Db	777	VGSBYVSRLIGLICITSTVOLVQLMPYGCILDHVENRGLGSDLLNMCQIAKGNSTYL	836
Oy	835	EDRLVHRDLAENNVLVKSPNHVKIIDFGLALDIDETVHAOGKVPKIMMALESILR	894
Db	837	EDRLVHRDLAENNVLVKSPNHVKIIDFGLALDIDETVHAOGKVPKIMMALESILR	896
Oy	895	RRLTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMV	954
Db	897	RRLTHOSDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMV	956
Oy	955	KCMWIDECGRPRERELVEFSRMAPQSFVVIQNEDEGPAISPDSPTSYSLIEDDDMGD	101
Db	957	KCMWIDECGRPRERELVEFSRMAPQSFVVIQNEDEGPAISPDSPTSYSLIEDDDMGD	101
Oy	1015	LVDAEYLVPOGCFPCDDPAPGAGCVHHRHRSSTRSGGDDLTGLPSESEAPRSPILA	107
Db	1017	LVDAEYLVPOGCFPCDDPAPGAGCVHHRHRSSTRSGGDDLTGLPSESEAPRSPILA	107
Oy	1075	PSBAGSDVDGDLGMAKGLQSLPTHPSPLQYSSDPIVPLPSBTDGVAPLITQSPQ	113
Db	1077	PSBAGSDVDGDLGMAKGLQSLPTHPSPLQYSSDPIVPLPSBTDGVAPLITQSPQ	113
Oy	1135	PEYVNOVDVAPQPPSPREGPLPAARPAGATLERAKTISPGKNGVVKOVFAFGAVENPEY	119
Db	1137	PEYVNOVDVAPQPPSPREGPLPAARPAGATLERAKTISPGKNGVVKOVFAFGAVENPEY	119
Oy	1195	LTPQGGAAPQHPHPAPFSPAFDNLYVMODDPERKAPSTFKGTPATENPEYGLADVPI	1253
Db	1197	LTPQGGAAPQHPHPAPFSPAFDNLYVMODDPERKAPSTFKGTPATENPEYGLADVPI	1255

```

RESULT 2
ERRB2 RAT
ID ERRB2 RAT STANDARD; PRT; 1257 AA.
AC P06494;
DT 01-JAN-1968 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (epidermal growth factor
DE receptor-related protein).
GN ERB2 OR NEU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma;
RX MEDLINE=86118662; PubMed=3945311;
RA Bargmann C.I., Hung M.-C., Weinberg R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related
RT protein.";
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lal C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially
RT expressed in the vertebrate nervous system.";
RL Neuron 6:691-704(1991).
RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Lotts F.J., Doak D.G., Mulvey D.,

```

```

RA Neman R,Cumpton M.J., Sternberg M.J.E., Campbell I.D.;  

RT "three dimensional structure of the transmembrane region of the proto-  

RL oncogenic and oncogenic forms of the neu protein.";  

EMBO J. 11:43-48(1992).  

CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  

CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GEF3 IS A  

CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  

CC ALPHA AND AMPHIREGULIN.  

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  

CC tyrosine phosphate.  

CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.  

CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.  

CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  

CC -1- PPM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  

CC RESIDUES (BY SIMILARITY).  

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  

-----  

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  

CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  

CC the European Bioinformatics Institute. There are no restrictions on its  

CC use by non-profit institutions as long as its content is in no way  

CC modified and this statement is not removed. Usage by and for commercial  

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  

CC or send an email to license@isb-sib.ch).  

CC -----  

DR EMBL; X03362; CAA27059.1; ALT_INIT.  

DR PIR; A24562; TVRTNU.  

DR HSSP; P11362; IRGU.  

DR InterPro; IPR0000494; EGFR_L_domain.  

DR InterPro; IPR000719; Euk_Pkinase.  

DR InterPro; IPR002174; Furin-like.  

DR InterPro; IPR001245; Tyr_Pkinase.  

DR InterPro; IPR004019; YLP motif.  

DR Pfam; PF00069; Pkinase; 1.  

DR Pfam; PF00757; Furin-like; 1.  

DR Pfam; PF01030; Recep_L_domain; 2.  

DR Pfam; PF02757; YLP_2.  

DR ProDom; PD000001; Euk_Pkinase; 1.  

DR SMART; SM00261; FU; 3.  

DR SMART; SM00219; TyKc; 1.  

DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.  

DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.  

DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.  

KM Transmembran; Glycoprotein; Multigene family; Receptor; Signal;  

KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  

KW Proto-oncogene; Disease mutation.  

FT SIGNAL 1 21  

FT CHAIN 22 1257  

FT DOMAIN 22 654  

FT TRANSMEM 655 677  

FT DOMAIN 159 369  

FT DOMAIN 473 646  

FT DOMAIN 722 989  

FT NP_BIND 728 736  

FT BINDING 755 755  

FT ACT_SITE 847 847  

FT DISULFID 196 205  

FT DISULFID 200 213  

FT DISULFID 221 228  

FT DISULFID 225 236  

FT DISULFID 237 245  

FT DISULFID 241 253  

FT DISULFID 256 263  

FT DISULFID 269 296  

FT DISULFID 300 312  

FT DISULFID 316 332  

FT DISULFID 335 339  

FT DISULFID 513 522  

FT DISULFID 517 530  

FT DISULFID 533 542  

FT DISULFID 546 562  

FT DISULFID 565 578  


```

FT DISULFID 569 586 BY SIMILARITY.  
 FT DISULFID 589 598 BY SIMILARITY.  
 FT DISULFID 602 625 BY SIMILARITY.  
 FT DISULFID 628 636 BY SIMILARITY.  
 FT DISULFID 632 644 BY SIMILARITY.  
 FT MOD\_RRS 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RRS 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).  
 SQ SEQUENCE 1257 AA, 138831 MW, 6129264583011402 CRC64;  
 Query Match 85.4%; Score 5817; DB 1; Length 1257;  
 Best Local Similarity 85.2%; Pred. No. 5,1e-306;  
 Matches 1075; Conservative 57; Mismatches 117; Indels 12; Gaps 4;

QY 1 MELALCRWGLLALIPGGAASCTGCTDMKRLPASPETHMLMHLVQCCVQGNL 60  
 DB 1 MELAMCRWGLLALIPGGAASCTGCTDMKRLPASPETHMLMHLVQCCVQGNL 60  
 QY 61 ELTYLPTNASLSPLODIOEVQYVLIANOVQVPLQRLIVRGTOLEFEDVYALAVLNDG 120  
 DB 61 ELTYLPANASLSPLODIOEVQYVLIANOVQVPLQRLIVRGTOLEFEDVYALAVLNDR 120  
 QY 121 DPLNNTPTVT-GASFGGLRELUASLSTILKGVLIQNPOLCYQDTILMKDIFHKNNOL 179  
 DB 121 DPONVAASTPGRTPEGRLQRLSTELILKGVLIQNPOLCYQDVMWLMADVFRKNOL 180  
 QY 180 ALLTIDTRSPACHPCSPMKGRCWGSSEDCOSLFTVQAGCAGACCKGPLPDCCHEQ 239  
 DB 181 APVIDITNRSRACPCACACCKDNHMGSPEDCOLITGICTSCACACCKGPLPDCCHEQ 240  
 QY 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMFNPBGRYTFGASCYTAC 299  
 DB 241 CAAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMFNPBGRYTFGASCYTTC 300  
 QY 300 PYNVISTDVGSCCTIVCPHNOEVTAEADTQCEKSKPCAVCYGLGMQVYKANSKFIGI 359  
 DB 301 PYNVISTEYVSCCTIVCPHNOEVTAEADTQCEKSKPCAVCYGLGMHHRGARALTSD 360  
 QY 360 TELEFAGCKKIFGSLAFIPESFDGPASNTAPLOPELOVETLEITGYLYISAMPDSL 419  
 DB 361 NVQEFDCCKIFGSLAFIPESFDGPSSGIAPLRPELOVETLEITGYLYISAMPDSL 420  
 QY 420 PDLVYFQNLQVIRGILHNGAYSLTLOGIGTSMGLRSLRELGSGLAIHHNTHLCFVHT 479  
 DB 421 RDLVYFQNLRIIRGILHNGAYSLTLOGIGTSHSLRSLRELGSGLAIHHRNAHLCFVHT 480  
 QY 480 VPMQDLFRNPHQALHTANRPEDE-CVGEGLACHOLCARGHGWGPTQVCNCSQFLRQ 538  
 DB 481 VPMQDLFRNPHQALHTANRPEDE-CVSSGLVCHNSLCAHGHGWGPTQVCNCSHFLRQ 540  
 QY 539 ECVESCRVLOGLPREYVNAARHCLPCHPECOQNGSVTQFGEADQVCAHAYKPPFYVA 598  
 DB 541 ECVESCRVWKGIPREYVNDKRLCPCHPECOQNPNSSETFGSEADQCAHAYKSSSCVA 600  
 QY 599 RCPGSKPDLSTYMPKFPDEEGACQPCPINTHSCVDLJDKGCAEORASPLTSISAV 658  
 DB 601 RCPGSKPDLSTYMPKFPDEEGACQPCPINTHSCVDLJDKGCAEORASPLTSISAV 660  
 QY 659 VGILLVVLGVFGILLI-----FNNFTVSEFWLRPKYSASLLEPLTSSGAMPQACMRI 712  
 DB 661 VGILLVVLGVFGILLI-----FNNFTVSEFWLRPKYSASLLEPLTSSGAMPQACMRI 716  
 QY 713 LKETELRKVVLGSGAFGVYKGINI PDGENKIVAIKVLRENTSPANKELIDEAYVM 772  
 DB 717 LKETELRKVVLGSGAFGVYKGINI PDGENKIVAIKVLRENTSPANKELIDEAYVM 776  
 QY 773 AGVSPYVSRLLGICLTSTVQLVTQMLPGCLLDHVRNKGRLSGQDLINMCQIAKMS 832

DB 777 AGVSPYVSRLLGICLTSTVQLVTQMLPGCLLDHVRNKGRLSGQDLINMCQIAKMS 836  
 QY 833 YLEDVRLVARDLAARNVLYKSPNHVKTDFGLARLLDIDETFEYHADGKVPYKMALESI 892  
 DB 837 YLEDVRLVARDLAARNVLYKSPNHVKTDFGLARLLDIDETFEYHADGKVPYKMALESI 896  
 QY 893 LRRRPHOSDWSYGTVTWELMTFGAKPVDGI-PAREIPDLLEKGRLLPQPICTIDVYMI 952  
 DB 897 LRRRPHOSDWSYGTVTWELMTFGAKPVDGI-PAREIPDLLEKGRLLPQPICTIDVYMI 956  
 QY 953 MYKCMWIDSECRPRFRELVSSESRMARDPQRFVVIQNEIDGASPLDSTFYRSLLEDDM 1012  
 DB 957 MYKCMWIDSECRPRFRELVSSESRMARDPQRFVVIQNEIDGASPLDSTFYRSLLEDDM 1016  
 QY 1013 GDLVDAEYLVYQQGFCDPAPGAGMYHHHRRSSSTSGGGDLTLGLEPSEBEERSP 1072  
 DB 1017 GDLVDAEYLVYQQGFCDPAPGAGMYHHHRRSSSTSGGGDLTLGLEPSEBEERSP 1076  
 QY 1073 LAPSEAGSDVDFDGLGMAAGLQSLPFDHDSPLQRYSEDEPTVPLPSETDGVAPLTC 1132  
 DB 1077 LAPSEAGSDVDFDGLGMAAGLQSLPFDHDSPLQRYSEDEPTVPLPSETDGVAPLTC 1136  
 QY 1133 PQPEYVNPDPVPRQPPSPREGPLPAARPAATLERAKTSLSPKNGVYKDVAFGAVENP 1192  
 DB 1137 PQPEYVNPDPVPRQPPSPREGPLPAARPAATLERAKTSLSPKNGVYKDVAFGAVENP 1196  
 QY 1193 EYLTPOGGAAPQHPAPSPAPFNDLYYWDOPPEPRGAPSPFKGPTANPEYGLDVP 1252  
 DB 1197 EYLTPOGGAAPQHPAPSPAPFNDLYYWDOPPEPRGAPSPFKGPTANPEYGLDVP 1256  
 QY 1253 V 1253  
 DB 1257 V 1257

RESULT 3  
 ERB2\_MESAU STANDARD; PRT; 1254 AA.  
 ID ERB2\_MESAU  
 AC 06053;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
 DE (p185erbB2) (NEU proto-oncogene) [C-erbB-2].  
 GN ERBB2 OR NEU.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 CX NCBI\_TaxID=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Nerve;  
 RX MEDLINE=94193007; PubMed=7908275;  
 RA Nakamura T., Uehijima T., Ishizaka Y., Nagao M., Arai M.,  
 RA Yamazaki Y., Ishikawa T.,  
 RT Cloning and activation of the Syrian hamster neu proto-oncogene.";  
 RL Gene 140:251-255(1994).  
 CC - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,  
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).  
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC - SUBCELLULAR LOCATION: TYPE I membrane protein.  
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

-----  
 DR EMBL: D16295; BAA03801.1; -  
 DR HSSP: P11362; IFCK  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR InterPro: IPR004019; YLP motif.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00261; FU; 3.  
 DR SMART: SM00219; TYKc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferrase; Tyrosine-protein Kinase; ATP-binding; Phosphorylation;  
 KW Proto-oncogene; Disease mutation.  
 FT SIGNAL 1 21  
 FT CHAIN 22 1254  
 FT DOMAIN 22 652  
 FT TRANSMEM 653 675  
 FT DOMAIN 676 1254  
 FT DOMAIN 158 368  
 FT DOMAIN 472 644  
 FT NP\_BIND 726 987  
 FT BINDING 753 753  
 FT ACT\_SITE 845 845  
 FT DISULFID 195 204  
 FT DISULFID 199 212  
 FT DISULFID 236 244  
 FT DISULFID 240 252  
 FT DISULFID 255 264  
 FT DISULFID 268 295  
 FT DISULFID 299 311  
 FT DISULFID 315 331  
 FT DISULFID 334 338  
 FT DISULFID 511 520  
 FT DISULFID 515 528  
 FT DISULFID 531 540  
 FT DISULFID 544 560  
 FT DISULFID 563 576  
 FT DISULFID 567 584  
 FT DISULFID 587 596  
 FT DISULFID 600 623  
 FT DISULFID 626 634  
 FT DISULFID 630 642  
 FT MOD\_RES 1139 1139  
 FT MOD\_RES 1247 1247  
 FT CARBOHYD 68 68  
 FT CARBOHYD 125 125  
 FT CARBOHYD 187 187  
 FT CARBOHYD 259 259  
 FT CARBOHYD 530 530  
 FT CARBOHYD 571 571  
 FT CARBOHYD 629 629  
 FT VARIANT 658 658  
 FT VARIANT 659 659  
 FT SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BE1 CRC64;  
 Query Match 85.1%; Score 5801.5; DB 1; Length 1254;  
 Best Local Similarity 84.8%; Pred. No. 3.5e-105;  
 Matches 1068; Conservative 66; Mismatches 114; Indels 11; Gaps 3;

QY	1	MELAAALRWGLLLALPPGAAS	TOVGTGTDPKRLTPAS	PETHLMLRHLYGCGVQGNL	60
DB	1	MELAAALRWGLLLALPPGAAS	TOVGTGTDPKRLTPAS	PETHLMLRHLYGCGVQGNL	60
QY	61	ELTYLPTNASISFLQDIQEV	GVYVLIANQVQVPLQRL	IVRGTQLFEDNYALALVQNG	120
DB	61	ELTYLPTNASISFLQDIQEV	GVYVLIANQVQVPLQRL	IVRGTQLFEDNYALALVQNG	120
QY	121	DLNNTPPVGASRGSLRELQ	LSLETLLKGVVLIQNNPQ	CVQDITLMQDIFKNNQLA	180
DB	121	DLNNTPPVGASRGSLRELQ	LSLETLLKGVVLIQNNPQ	CVQDITLMQDIFKNNQLA	180
QY	181	LTLIDTRSRACHPCEPMCK	SRQCESSSEDCQSLTR	TVACGACGARKGRLPTDCHEQC	240
DB	181	LTLIDTRSRACHPCEPMCK	SRQCESSSEDCQSLTR	TVACGACGARKGRLPTDCHEQC	240
QY	241	AACTGPKHSDCLALFPHNS	GTIELHCPALVYNTNT	TFESMPNPEGRYTTGACCVTACP	300
DB	241	AACTGPKHSDCLALFPHNS	GTIELHCPALVYNTNT	TFESMPNPEGRYTTGACCVTACP	300
QY	301	YNYLSTVGSCTVCPVHNO	EVTAEDGTQRCCKSKPC	ARCYGLGMQYIKANSKFIGIT	360
DB	301	YNYLSTVGSCTVCPVHNO	EVTAEDGTQRCCKSKPC	ARCYGLGMQYIKANSKFIGIT	360
QY	361	ELERAGCKKIFGSLAFPE	SFSDGDPASNTAPLQ	EQLOVETLEITGYLISAMPDLP	420
DB	361	ELERAGCKKIFGSLAFPE	SFSDGDPASNTAPLQ	EQLOVETLEITGYLISAMPDLP	420
QY	421	DLSPFQNRVIRGVVLDG	AYSIALQIGIRMLGR	LSRSLREGSSVLIRHRTLCFVHTV	480
DB	421	DLSPFQNRVIRGVVLDG	AYSIALQIGIRMLGR	LSRSLREGSSVLIRHRTLCFVHTV	480
QY	481	PMDFLFRPHQALHTANR	PEDECVGGLACHOL	CARGHCWGPPTQCVNCSQPLRQEC	540
DB	481	PMDFLFRPHQALHTANR	PEDECVGGLACHOL	CARGHCWGPPTQCVNCSQPLRQEC	540
QY	541	VEEGRVLDGLFREYVNA	RHCLPCHPEQOPQNG	SVTTCGPENQCVACAHYDPPFCVARC	600
DB	541	VEEGRVLDGLFREYVNA	RHCLPCHPEQOPQNG	SVTTCGPENQCVACAHYDPPFCVARC	600
QY	601	PSGVKPDLSYMPIKFP	DEEGACQPCPINC	THSCVDDDKQCPAEOBAPLTSIVSAVG	660
DB	601	PSGVKPDLSYMPIKFP	DEEGACQPCPINC	THSCVDDDKQCPAEOBAPLTSIVSAVG	660
QY	661	ILLVVYVGVGCIIL	-----FNNFTYS	FMLRVRKVSASHLEPLTPSGAMPNQOMRLK	714
DB	661	ILLVVYVGVGCIIL	-----FNNFTYS	FMLRVRKVSASHLEPLTPSGAMPNQOMRLK	714
QY	715	ETELRKVYVLSGAF	GTYYKGIWIPDGEN	VYIPIAIVKYLRENTSPKANKELDFAVYVAG	774
DB	715	ETELRKVYVLSGAF	GTYYKGIWIPDGEN	VYIPIAIVKYLRENTSPKANKELDFAVYVAG	774
QY	775	VGSPVYSLIGICL	TSVQVLTQVLTQ	MPYGCILLDHYRENGRGLSODLLNMCQIAKMSYL	834
DB	775	VGSPVYSLIGICL	TSVQVLTQVLTQ	MPYGCILLDHYRENGRGLSODLLNMCQIAKMSYL	834
QY	835	EDVLYVHEDLAARV	LVYSPPHVKITP	PGCLARLLDIDETEHNAQGVKPIFMALFESTLR	894
DB	835	EDVLYVHEDLAARV	LVYSPPHVKITP	PGCLARLLDIDETEHNAQGVKPIFMALFESTLR	894
QY	895	RRFTQSDVMSY	GVTVLMETFGAKPY	DGIPAREIPDLLEKGERLPPPCITIDVYIMV	954
DB	895	RRFTQSDVMSY	GVTVLMETFGAKPY	DGIPAREIPDLLEKGERLPPPCITIDVYIMV	954
QY	955	KCMWIDSECR	RFRELVSEFSR	MRADPQRFVVIQNE	DGLSPASPLDSTFYRLLDDDDMGD
DB	955	KCMWIDSECR	RFRELVSEFSR	MRADPQRFVVIQNE	DGLSPASPLDSTFYRLLDDDDMGD
QY	1015	LVDAEYVLPQ	CGFFCPDPAPAG	AGVYHRRSSSTRSGCGDLTLGLPSESEAPRSPLA	1074
DB	1015	LVDAEYVLPQ	CGFFCPDPAPAG	AGVYHRRSSSTRSGCGDLTLGLPSESEAPRSPLA	1074
QY	1076	LVDAEYVLPQ	CGFFCPDPAPAG	AGVYHRRSSSTRSGCGDLTLGLPSESEAPRSPLA	1076
DB	1076	LVDAEYVLPQ	CGFFCPDPAPAG	AGVYHRRSSSTRSGCGDLTLGLPSESEAPRSPLA	1076

QY 1075 PSEAGSDVFDGDIQWGAAGLQSLFTHDPSPLQVRSSEDTVPJPSSTDCYVAPLTCSPQ 1134  
 DB 1077 PSEAGSDVFEGETIGMATGKATGPOSSISPRDSSPLQVRSSEDTLPJFTDTCYVAPLACSPQ 1136  
 QY 1135 PEYNOGPDVPPQPSPEPGLPAPARPAGATLERAKTSSPKNGVAKVQVAFGGAVENPEY 1194  
 DB 1137 PEYNOGPDVPPQPSPEPGLPAPARPAGATLERAKTSSPKNGVAKVQVAFGGAVENPEY 1196  
 QY 1195 LTPOGGAPOPHPPAPASPAFDNLVYWDQPPERGADEPSTFKGTPTAENPEYLGDDVPV 1253  
 DB 1197 LVPRGASGQPH-PPALCPAFDNLVYWDQPPERGADEPSTFKGTPTAENPEYLGDDVPV 1254  
 RESULT 4  
 EGFR\_HUMAN STANDARD; PRT; 1210 AA.  
 ID EGFR\_HUMAN  
 AC P00533; P06268; Q14225; Q9UMD7; Q9JMD8; Q9UMG5; Q92795; O00732;  
 AC O00688; Q9H252; Q9H2C9; Q9GZX1; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor  
 DE protein-tyrosine kinase EDB-1).  
 GN EGFR OR ERBB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=84219729; PubMed=6328312;  
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,  
 RA Mayes E.L.V., Mittelman N., Waterfield M.D., Seeburg P.H.;  
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant  
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";  
 RL Nature 309:418-425(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE=95382957; PubMed=7654368;  
 RA Tlekis J.V., Stark B.C., Scoccia B.;  
 RT "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta.";  
 RL Mol. Reprod. Dev. 41:149-156(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE=97076686; PubMed=8918811;  
 RA Reiter J.L., Mainle N.U.;  
 RT "A 1.8 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor.";  
 RL Nucleic Acids Res. 24:4050-4056(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX MEDLINE=97256547; PubMed=9103388;  
 RA Tlekis J.V., Garitti J., Niederberger C., Scoccia B.;  
 RT "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGFR) in ovarian cancer.";  
 RL Gynecol. Oncol. 65:36-41(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RX MEDLINE=21100872; PubMed=1161793;  
 RA Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,  
 RA Scheel Sincilar C., Pearall R.S., Green P.J., Yee D., Lampfand A.L.,  
 RA Balasubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,  
 RA Mainle N.U.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGFR transcripts encoding truncated receptor  
 RT isoforms.";  
 RL Genomics 71:1-20(2001).  
 RN [6]  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6320621;  
 RA Lin C.R., Chen W.S., Krulger W., Stolarsky L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
 RT "Expression cloning of human EGF receptor complementary DNA: gene  
 RT amplification and three related messenger RNA products in A431  
 RT cells.";  
 RL Science 224:843-848(1984).  
 RN [8]  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.;  
 RT "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAs overproduced in A431 carcinoma cells.";  
 RL Nature 309:806-810(1984).  
 RN [9]  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Haley J.D., Whittle N., Bennett P., Kitchington D., Ullrich A.,  
 RA Waterfield M.D.;  
 RT "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription.";  
 RL Oncogene Rev. 1:375-396(1987).  
 RN [11]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=9107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RT "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RT factor receptor proto-oncogene RNA synthesis.";  
 RL J. Biol. Chem. 266:1746-1753(1991).  
 RN [12]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RT "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN [13]  
 RP SEQUENCE OF 540.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honnegger A.M.,  
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 RT



RT receptor and HER2/neu are located in their carboxyl-terminal tails.  
 RT Identification of a novel site in EGF receptor."  
 RL J. Biol. Chem. 264:10667-10671(1989).  
 RN [16]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RX MEDLINE=96398132; Pubmed=8962717;  
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.,  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 RT the epidermal growth factor receptor expressed in Chinese hamster  
 RT ovary fibroblasts."  
 RL Growth Factors 13:121-132(1996).  
 RN [17]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; Pubmed=10731668;  
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.:  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor  
 RT receptor."  
 RL J. Biochem. 127:65-72(2000).  
 RN [18]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=98225196; Pubmed=9556602;  
 RA Abe Y., Odaka M., Inagaki F., Iax I., Schlessinger J., Kohda D.:  
 RT "Disulfide bond structure of human epidermal growth factor receptor."  
 RL J. Biol. Chem. 273:11150-11157(1998).  
 RN [19]  
 RP REVIEW.  
 RX MEDLINE=87297456; Pubmed=3039909;  
 RA Carpenter G.:  
 RT "Receptors for epidermal growth factor and other polypeptide  
 RT mitogens."  
 RL Annu. Rev. Biochem. 56:881-914(1987).  
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF  
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is  
 CC involved in the control of cell growth and differentiation.  
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 CC secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms: 1/p170 (shown here), 2/p60/  
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by  
 CC alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 CC expressed in ovarian cancers.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC  
 CC EMBL; X00588; CA25240.1; -  
 CC EMBL; U05089; AAB53063.1; -  
 CC EMBL; U48722; AAC50802.1; -  
 CC EMBL; U48723; AAC50804.1; -  
 CC EMBL; U48724; AAC50796.1; -  
 CC EMBL; U48725; AAC50797.1; -  
 CC EMBL; U48726; AAC50798.1; -

Query Match 45.0%; Score 3067; DB 1; Length 1210;  
 Best Local Similarity 49.1%; Pred. No. 6.3e-158;  
 Matches 621; Conservative 174; Mismatches 362; Indels 108; Gaps 24;

QY 11 LLLALLPGEAA--STVCTGTDMKRLPASBETHLDMRHLYOGCVVQGNLELYLPTN 68  
 DB LLLAALCPASRLAEKKVCGTSNKLTOGLTFEDHLSLQRMNNECVLLGNLEITYQRN 73  
 QY 69 ASLSFLDIOEVOGVYLLAHNOVRQVPLQRLRYVGTQGFEDNYLAVLUNGCDPLNNTPE 128  
 DB YDLSFLKTIQEVAGVYLLALNVTVERIPLENLIQIIGNMYENSVALAVLSND----- 126  
 QY 122 VTGASPGRLRELOLSLREILKGGVLIQRNPOLQODITLIMMDIHKNNQALLTLIDTR 188  
 DB ---ANKTKLELPMNLQELILGAVRFSNNPLCWVESIQMFDIYSSDLSMMSMDPQNH 183  
 QY 189 SRACHPCSEPMCKGRSBCWSESSDCSLTRTVAGGCA-RCKGPLPTDCCEQCAAGCTGP 247  
 DB LGSQCKDPSRCNGSCWNGEENCCQKLRITCAQCCSRRCRCKSPSCDCHNCAAGCTGP 243  
 QY 248 KNSDCLACHFNHSGICELHCFALVTYNTDTRESNPNREGRTFGASCTYACPPYLYSD 307  
 DB RESDCLVCRKRFDEATCDTCPLMLYPTTYQMDVNPBGKYSFGATCVKCPRYVYVTD 303  
 QY 308 VGSCTLVCPLNHQTAEADGTQRCCEKCKPCARVCYGMQYIKANSKFIGITEDE-FAG 366  
 DB HSCYRACGADSYEM-EEDGVKCKCKCGCPCKVNGNGIGEFK-DLSLNATNPKHFQN 361  
 QY 367 CKKIFGSLAFLEPSFDGDPASNTAPLOEQLOVFTLEITGYLYISAPDSLPDLSYFQ 426  
 DB CTSISGDHLILFVAFRGDSFTHTPDLDEQDILITKVEILNGFLIQAMPENRDLHAFE 421  
 QY 427 NLOVRGRILHNGAASLTLOGISLMLGRSLREISGLALHNHTHLCFYTVTMDOLE 486  
 DB NLEITRGKTKQKQGSFLAVSNLNTSLGRSLKESIDGVDVITSGNKLQYATNTWKKLF 481  
 QY 487 RNPQALLHTANRPDECEVGEGLACHQDLCARGHCWPGPTQVCNCSQPLRSGOCEVECRV 546  
 DB GTSQCKTKILSNRGNSCATQGVCHALCSPEGCWGPBRDVCSSRNNSRGRECVCKNL 541  
 QY 547 LOGLREYVNAHCLPCHPECCPQNGSVTCFSPENDQCVACHYKADPPCVARCSGVKP 606  
 DB LGEGRREFENSECIQCHPECLPQAMNITTCRGPDNCTIQCHYIDGPVCVTCAGVWG 601  
 QY 607 DLSYPMKFPDEBACQCPPINCHSCVDLDDKCPAQRASPLTSIYSAVG--IIL 663  
 DB ENNTL-VMKYAAAGVCHLCHNCYTCGCPGLBECPNTNGRPKIP--STATNVGALLLL 658  
 QY 664 VVLGVVSGILIFNNPTV-SFWLAVPKVSASHLEPLTPSGAMPNOQMILKETELRYK 722  
 DB VVALGI--GLFMRHRHIVKRTLRLLQERLEVEPLTPSGEAPNOALLRIKETEFKKIK 716  
 QY 723 VGSAGFVTVYKGIWIPDENVKIPVAIKVIRENTPSPKXKILDEAYVMAGVSPYSR 782  
 DB VGSAGFVTVYKGIWIPDEKKTIVALKELREATSPKXKILDEAYVMASVDNPHVR 776  
 QY 783 ILGICLTSTVQVLTQIMPYGLLDHVRNRLGSGDLLNMCQIAKGSYLEDEVLYHR 842  
 DB LIGICLTSTVQVLTQIMPFGCLLDYREKNDIGSYLLNMCVQIAKGNVYLEDRLVHR 836  
 QY 843 DLAANVLYKSNHKKITDFGLARLDIDETFNADGKVPKMMALSLIRRTTHSD 902  
 DB DLAANVLYKTHQHKITDFGLAKLGAEEKYHAEGGVPIKMMALSLIRRTTHSD 896  
 QY 903 VMSYGVTVLWELTFGAKPYDGIIPAREIDPLEKGRLLPOPICTIDYVMIVKCMMDISE 962  
 DB VMSYGVTVLWELTFGSKPYDGIIPASEISILEKGRLLQPICTIDYVMIVKCMMDIAD 956  
 QY 963 GPRFRELVSFSKMAKDPQRFVTVQ-NEDLCFAPPLDSTFRSLLEDNDDGLVDAEY 1021  
 DB SRKPRELILFESKMAKDPQRFVTVQGERNMLPSPPTSNFPRALMDEEDMDVDVDAEY 1016  
 QY 1022 LVPOGFFPCPDPAAGVGVHHRHRSSTRSGGDLTLGLESESEAPRSPLAPEGAGS 1081  
 DB LVPOGFF-----SPTSTRTPLSLSLATS 1042

```

QY 1082 DVEDGDIQMGAAKGLQLPHDPSLPLOYSSEDTVPVLPST--DGVIAPLTGSPQPEYVN 1139
D 1043 N--NSTVACIDRNGLQSCPIEDSFLOYSDDPGALTEDSIDDTFL-----PVPEYIN 1094
QY 1140 QPDVPPQPSPREPPLAARPAAGTLEPARKTLSPGKGVYKDVAFAGAVENPEYL-TPQ 1198
D 1095 Q-SVFKRPPAGSVQVPPVHNQDLPN-----ASRPPHQD--PHSTAGNPEYLVNQ 1143
QY 1199 GGAAPQHPPPAFSPAFDNLVYWDQ-----DP-----PERGAPSTPKGTPTAE 1242
D 1144 -----PTCVNSTFSDSPAHMAQKSHQISLDNPDYQDFFPKAKPNGIFKGS-TAE 1193
QY 1243 NPEYL 1247
D 1194 NAEYL 1198

RESULT 5
EGFR_MOUSE
ID EGFR_MOUSE STANDARD: PRT: 1210 AA.
AC Q01279
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112).
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=liver;
RX MEDLINE=93026370; PubMed=1408137;
RA Avivi A., Skorecki K., Yayon A., Givol D.,
RT "Promoter region of the murine fibroblast growth factor receptor 2
RT (bex/KGFR) gene.";
RL Oncogene 7:1957-1962 (1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; and CD-1; TISSUE=liver, and Decidua;
RX MEDLINE=93126380; PubMed=7678348;
RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
RT "Expression of the epidermal growth factor receptor gene is regulated
RT in mouse blastocysts during delayed implantation.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59 (1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=liver;
RA Hibbs M.L.;
RT Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=B6/C3; TISSUE=Liver;
RX MEDLINE=94170986; PubMed=812525;
RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Baird H.S.,
RT "The mouse waved-2 phenotype results from a point mutation in the EGF
RT receptor tyrosine kinase.";
RL Genes Dev. 8:399-413 (1994).
RN [5]
RP SEQUENCE OF 1-714 FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=91232866; PubMed=2030916;
RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
RT "Comparison of EGF receptor sequences as a guide to study the ligand
RT binding site.";
RL Oncogene 6:673-676 (1991).
RN [6]
RP SEQUENCE OF 969-1117 FROM N.A.
RC STRAIN=C3H;
RA Bisinger D.P., Serrero G.;
RT Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.

```

---

```

CC CC - FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC CC tyrosine phosphate.
CC CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC CC - MISCELLANEOUS: Binding of EGF to the receptor leads to
CC CC dimerization, internalization of the EGF-receptor complex,
CC CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC CC synthesis, and cell proliferation.
CC CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
DR EMBL: X78987; CAA55587.1; -
DR EMBL: U03425; AAA17899.1; -
DR EMBL: X59698; CAA42219.1; -
DR EMBL: L06864; AAA53029.1; -
DR EMBL: Z12608; CAA78249.1; -
DR HSSP: P11362; 1FGK.
DR MGD: MGI:95294; Egfr.
DR InterPro: IPR000484; EGFR_L_domain.
DR InterPro: IPR000719; Euk_Pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_Pkinase.
DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recept_L_domain; 2.
DR ProDom: PD000001; Euk_Pkinase; 1.
DR SMART: SM00261; Fur_3.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR TransMembrane: Glycoprotein: Receptor; Signal; Transferase;
KM Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
FT SIGNAL 1 24
FT CHAIN 25 1210
FT DOMAIN 25 647
FT TRANSMEM 648 670
FT DOMAIN 671 1210
FT REPEAT 75 300
FT REPEAT 390 600
FT DOMAIN 1028 1071
FT DOMAIN 714 981
FT NP_BIND 720 728
FT BINDING 747 747
FT ACT_SITE 839 839
FT DISULFID 190 199
FT DISULFID 251 260
FT DISULFID 264 291
FT DISULFID 295 307
FT DISULFID 311 326
FT DISULFID 329 333
FT DISULFID 506 515
FT DISULFID 510 523
FT DISULFID 526 535
FT DISULFID 539 555
FT DISULFID 558 571
FT DISULFID 562 579
FT DISULFID 582 591
FT DISULFID 595 617
FT DISULFID 595 617
BY SIMILARITY.

```

FT DISULFID 620 628 BY SIMILARITY.  
 FT DISULFID 624 636 BY SIMILARITY.  
 FT MOD RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 FT MOD RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD RES 1197 1197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 128 128 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 175 175 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 352 352 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 413 413 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 444 444 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 528 528 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 568 568 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 603 603 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CONFLICT 19 19 C -> S (IN REF. 2).  
 FT CONFLICT 539 539 C -> W (IN REF. 5).  
 FT CONFLICT 991 991 L -> F (IN REF. 4).  
 FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).  
 SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46D2D2F5 CRC64;

Query Match 44.6%; Score 3041; DB 1; Length 1210;  
 Best Local Similarity 48.9%; Pred. No. 1,6e-156; Index 114; Gaps 25;  
 Matches 622; Conservative 168; Mismatches 369;

11 LLLALLPPGAA--STGYCTGTDMKRLPASPTHLDRLHLYOGCOVQGNLELYLPTN 68  
 14 LITLALCAAGALEBEKKYCGCTSNRLTGLTFEDHFLSLQRMNNCCEVLGNLEITYVQNN 73  
 69 AASLPLDIDIEVGVYLIANNQVRLQVLRVGRVQLFEDNYALAVDNGPLNNTTP 128  
 74 YDISFKTIEVGVYLIANNQVRLQVLRVGRVQLFEDNYALAVDNGPLNNTTP 124  
 129 VTGASPGGLRELALSLTEILKGVLIQRNPOCYOCTILMKDI---FHKNQALATLI 184  
 125 -YGTNRGLRELPMRNQGLILGAVRSNNPILCMNDTIOMRIYQNVFMSNML--- 180  
 185 DTNRSRACHPGSPCKSCSRGSSSEDDQSLRTVACAGCA-RCKRPLPTDCHEGCAAG 243  
 181 -QSHPSGCPKCDSPGSPGSGGSEENCQKLTIKICQOCSHRCRGRSPDCCNCAAG 239  
 244 CTGPKASDCLACIHFHNSGICELHCPALVTYNTDFESMNPGRGRTFGASCYACAPVY 303  
 240 CTGPRBBDCLVCCQKQDEATCTCTCPMLYNTTIOMDVNPGKTSFGATCYKCPRY 299  
 304 LSTDVGSCTLVCPILHNOEVTAEQGTORCEKSKPCARVYGLMOYIKANSKFIITEL 363  
 300 VVMDHSGCVACGPDYEV-EEDGIRKCKKCDGPRKVCNGIGIFEK-DTSLNNTNIK 357  
 364 -FAGCKKIFGSLAFSPDFGDASNAPLQPSQLOVETLEITGTYLISANPDSLPL 422  
 358 HFYKCTAISGDLHILPAFKGDSPTPTPLDPRELELTKVKTITFLLIQAAPDMWTL 417  
 423 SVFQNLQVIRGRILHNGAVSLTQGLISWLGRLSRELSSGALIHNTLGFVTPW 482  
 418 HAFENLEIRGRTRKHQGFSLAVGINTSLGRSLKEISDGVILISGRNLCYANTIM 477  
 483 DQFRNPHOALLTANRPEDECVGEGILACHQLCARHGCHWPGTQCVNCSQFLRGEV 542  
 478 KKLFTPNQTKIMNNAEKDCAVNHVNCPLSSSECCWPREPDCVSCNVSRGRECV 537  
 543 ECVLQGLPREYVNAHCLPCHCECPONGSVTCFGEADQVACAHYKDPFCVAPCS 602  
 538 KNILSEGEPRFVENSICICHECLPQANNTICTRGPNCTICQAHYIDGRPVATCA 597  
 603 GVKPDLSTYPMKFPDEGACQCPINCHSCVDLDDKGPAPQASPLTSYSAVGL 662  
 598 GINGENNNTL-VMKYADANNVCHCHANCITYGCAGPGIQQGEVWPSGPKPSIATIGVGL 656  
 663 LVVVLGVVFGILIF---NNFTVSFWLRVPKVSASHLEPLTPSGAMPNQAMRLKETEIR 719

DB 657 LFIIV-VALGIGLFRRRHIVKRTLRQLQRELEVLEPISGEAPNOAHILKETEER 715  
 QY 720 KYKULGSAFGVYVGMIPGSENVKIVAVIUVLEBNSPKANKELILEAVYMACVGSY 779  
 DB 716 KIKVIGSAGFVYVGLMIPSEKIKIPALILEBNSPKANKELILEAVYMACVGSY 775  
 QY 780 VSRLGICLITVQVLTQVLMIPYGLLDHVRNKGSLGSDLLNMCQIAKGSYLEDVL 839  
 DB 776 VCRLLGICLITVQVLTQVLMIPYGLLDHVRNKGSLGSDLLNMCQIAKGSYLEDVL 835  
 QY 840 VHRDLAARVLYKSNHYKITTDFGLARLLDIDETVHADGCVPIKMMALBETLRRFTH 899  
 DB 836 VHRDLAARVLYKSNHYKITTDFGLARLLDIDETVHADGCVPIKMMALBETLRRFTH 895  
 QY 900 QSDVMSYGVYVLEMTFGAKPYDGIIPAREIPLEKGERLPPICTIDVYIMYKMMI 959  
 DB 896 QSDVMSYGVYVLEMTFGAKPYDGIIPAREIPLEKGERLPPICTIDVYIMYKMMI 955  
 QY 960 DSECRPRRELVSSEFSRMAPDQRFVITQ-NEDLGASPLDSTFYRSILEDMDGLDVA 1018  
 DB 956 DADSRPKFRELILEFSKMARDPQRYLQIGDERMHLPEPDTSNFYRALMEDMEDVDVA 1015  
 QY 1019 EEYLPQCGFFCPDPAFGAGGVHHRHSSSTRSGGDLTLGLESESEAPRSPLAPSG 1078  
 DB 1016 DEYLLPQCGFF-----NSPST-----SKTPLLSSLS 1041  
 QY 1079 AGSDVFDGDLGMAAGLQSLPTHDSPLORYSEDPVLPSET--DGYVAPLTCSPPE 1136  
 DB 1042 ATSN-----NSTVACINRSGSCVKEKEDAFLORRSSDPTGAVYEDNIDDAFL-----PYDE 1091  
 QY 1137 YNQGDPVAPQPSPEGFLPAAPRAGATILERAKTLSFGKGVYKGVFAAGAVENPEYL 1195  
 DB 1092 YVNO-SVPRRPAQSVQNPVYHNPQPLP-----APGRDLHYN--PHSNVAGNPEYL 1140  
 QY 1196 TPQGAAPQPPHPPAPSPAFNLTLYWDQ-----DP-----PERGAPPSFTKGP 1239  
 DB 1141 TQO-----PTLSSGFSNPAIMQKSHQSLNDPDIYQODFPFKETKNGIFKG-P 1190  
 QY 1240 TAENPEYGLDVP 1252  
 DB 1191 TAENAEYLRVAP 1203

RESULT 6  
 ERB4\_HUMAN  
 ID ERB4\_HUMAN STANDARD; PRT; 1308 AA.  
 AC Q15303;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)  
 DE (p180erbB4) (tyrosine kinase-type cell surface receptor HER4).  
 GN ERB4 OR HER4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=93189574; PubMed=8383326;  
 RA Plozman G.D., Couloucou J.-M., Whitney G.S., Green J.M., Carlton G.W.,  
 Foy L., Neubauer M.G., Shoyab M.,  
 "Ligand-specific activation of HER4/p180erbB4, a fourth member of the  
 RT epidermal growth factor receptor family,"  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750 (1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM JM-A AND JM-B).  
 RC TISSUE=fetal brain;  
 RX MEDLINE=97476287; PubMed=9334263;  
 RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plozman G.D.,  
 Klagsbrun M.,

	FT	DISULFID	234	246	BY SIMILARITY.
	FT	DISULFID	249	258	BY SIMILARITY.
	FT	DISULFID	262	289	BY SIMILARITY.
	FT	DISULFID	293	304	BY SIMILARITY.
	FT	DISULFID	308	323	BY SIMILARITY.
	FT	DISULFID	326	330	BY SIMILARITY.
	FT	DISULFID	503	512	BY SIMILARITY.
	FT	DISULFID	507	520	BY SIMILARITY.
	FT	DISULFID	523	532	BY SIMILARITY.
	FT	DISULFID	536	552	BY SIMILARITY.
	FT	DISULFID	555	569	BY SIMILARITY.
	FT	DISULFID	559	577	BY SIMILARITY.
	FT	DISULFID	580	589	BY SIMILARITY.
	FT	DISULFID	593	614	BY SIMILARITY.
	FT	DISULFID	617	625	BY SIMILARITY.
	FT	DISULFID	621	633	BY SIMILARITY.
	FT	MOD_RES	1162	1162	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
	FT	MOD_RES	1188	1188	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
	FT	MOD_RES	1258	1258	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
	FT	MOD_RES	1284	1284	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
	FT	CARBONYD	138	138	N-LINKED (GLCNAC. .) (POTENTIAL).
	FT	CARBONYD	174	174	N-LINKED (GLCNAC. .) (POTENTIAL).
	FT	CARBONYD	181	181	N-LINKED (GLCNAC. .) (POTENTIAL).
	FT	CARBONYD	253	253	N-LINKED (GLCNAC. .) (POTENTIAL).
	FT	CARBONYD	358	358	N-LINKED (GLCNAC. .) (POTENTIAL).
	FT	CARBONYD	410	410	N-LINKED (GLCNAC. .) (POTENTIAL).
	FT	CARBONYD	473	473	N-LINKED (GLCNAC. .) (POTENTIAL).
	FT	CARBONYD	495	495	N-LINKED (GLCNAC. .) (POTENTIAL).
	FT	CARBONYD	548	548	N-LINKED (GLCNAC. .) (POTENTIAL).
	FT	CARBONYD	576	576	N-LINKED (GLCNAC. .) (POTENTIAL).
	FT	CARBONYD	620	620	N-LINKED (GLCNAC. .) (POTENTIAL).
	FT	VARSPLIC	626	648	NGTSHDCIYPTGTGSLRQHA -> IGSSIEDCIGLMD (IN ISOFORM JM-B).
SQ	SEQUENCE	1308 AA;	146807 MW;	5E4A80965D88761 CRC64;	
Query Match		42.7%;	Score 2908.5;	DB 1;	Length 1308;
Best Local Similarity		44.4%;	Pred. No. 2.3e-149;		
Matches 600;		Conservative 182;	Mismatches 384;	Indels 185;	Gaps 28;
QY	9	WGELLALLPPGAA----	STGVCTGTDMKRLPASPETHLDMLRYGGCCVVGCSNLETTY	64	
Db	8	WVVWSLLVAGTGVPDSGSVCAGTEKLSLSLDLBQYALRKYYENCEVMGNLEITS	67		
QY	65	LFPNLSISFLDIODIOEGYVLIANHYROYPLRLRVRGOLFEENDYVALAVLNDGPBLN	124		
Db	68	IENHRDLSFRSREVTGYVLALNQRYPLENRIRIKRTKLVEDRYALAIPIFYRKDG	127		
QY	125	NTTPVTVASBQGLEQLGRSLTELKSGVLIQRNPOLCYODTILMKDI FHKNNLATLI	184		
Db	128	NF-----GLQELGLKNLTLLINGVVDQKKFLCYADTIHMODIVRNPMPSNLTLV	178		
QY	185	DTRSRACHPCSPMCKSGRCMGESSEDDCSLTRVTCAGGC-ARGKGPLPTDCCHEOCAG	243		
Db	179	STNGSSCGCGCHKSCTG-RKMGPLENHCQTLLRTVLCABQCDGRCTGYGVSDSCCRBCAG	237		
QY	244	CTGPKHSDDLCLAFHNHSGICELHCPALVTYNDDTFESMPNREGRYTFGASCYTACPXYN	303		
Db	238	CSGSKXODCACMKFNDSGACVTCQPOTFFVYNPTTFLEHNFNAKYUYGAFCVVKCPDNF	297		
QY	304	LSTDVGSCTLVCPHLNHGEATAEDETOCEKSKXCPCAVCVGGLGMQYKANSKFIGITELE	363		
Db	298	V-VDSSCVYACPSBSKKEV-EENGDKMKCKECTDIPCAKDQSIGTSGLSMAQUTVDSSNIDX	355		
QY	364	FAGCKRIFGSLAPLPESFDGDPASNTAPLOPEOLOVEFTELEITGYLYISAMPDSLPLDS	423		
Db	356	FINCTKTINGMLIFLVTDIGHDPYNAIAIDPEKLNINVRTRYAREITGFNLINSWPNNMDFS	415		
QY	424	VFOUUVYIRGRLIHNGAYSLTLOGLGSMWGLRSLRELGSGDALIHNNTHLCFNYHPWD	483		
Db	416	VFSULVITIGRAVLVYSGLILLKQGTTLSLOFGSKAISAGNIYITDNSNLCTYHTTNMT	475		
QY	484	QLFENPHQAIIHTANRDEDECVEGGLACHOLCARGHGCMWRGPPTQCVNCSQFLRGQECVEE	543		

```

Db      476 TLFSTINQIVIRDNKKAENCTAEGWVGNHLCSSDGCWGPDPQCLSCRRFSRGRICTES 535
Qy      544 CAVLGLPREVYNNARHCLPCHPECCP-QNGSVTCGCPREDDQCAAHYNDPPFCARCS 602
Db      536 CNLYDEEFREFEFGSICVECDPQCEMEDGLTCHGPGPDNCTKCSHFADGNCVEKCPD 595
Qy      603 GVKPDLSTYPMIKWKFDEBACQPCPINCTHSCVDDDKCC-----PAEORASPL 651
Db      596 GAGGANSF--IFRYADPDRECHCHCHNCTGCGNPTSHDCIYYPMWGHSTLQGHAR-TPL 652
Qy      652 TSIVSAVY-GILLVVLGVVPGILLFNNFTVSFWLRVPVNSH-----LEPITPSGA 703
Db      653 --IAAGVIGLFLVGLTFLFAVYARR-----SIKKRLRRLFELETELVEPLTPSGT 703
Qy      704 MPEQAOMLIKETELRKVKVLSGAFYVKGIMWPDGENVIPAIVKLVRENTSPKANK 763
Db      704 ANQOQRLIKETELRKVKVLSGAFYVKGIMWPDGENVIPAIVKLVRENTSPKANK 763
Qy      764 EILDEAYVAVGSPYVSRLLGICLTSTVQVLTQMLPVGCLLDHYRENRGLSGODLLNW 823
Db      764 EEMDEALIMASMDHPHLVRLGLVCLSPITQVLTQMLPHGCLLEYVHEHKNIGSOLLNM 823
Qy      824 CMOIAGMAGYLEDVRLVHRDLAARVLYKSPHVKITDGLARLLDIDETEVHADGKVP 883
Db      824 CVOIAGMAGYLEERELVHRDLAARVLYKSPHVKITDGLARLLDIDETEVHADGKVP 883
Qy      884 IKMALESILRRRFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPP 943
Db      884 IKMALECIHYRKFTHQSDVWSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPP 943
Qy      944 ICTIDVYVMWCMWITDSECPREFRLVSEFSRMAADPQRYVIONED-LGASPLDSTF 1002
Db      944 ICTIDVYVMWCMWITDSECPREFRLVSEFSRMAADPQRYVIONED-LGASPLDSTF 1002
Qy      1003 YKSLLEDDMDGLVDAEEYLVPOQGFCCDPAPAGAGWVHRRSSSTRSGGDLTLGLE 1062
Db      1004 FQNLDEEDLEDDMDGEEYLV-PQANIPPP-----ITSRARIDSNKS-----EIGHS 1051
Qy      1063 PSEEEAPRS-----PLAP-SEGAGSVFPGDGLGMAA 1093
Db      1052 PPAYATPMGNGFVYRDGFAAEQGVSVYRAPTSTIPAPVAGATAIPIFDSCNGTL 1111
Qy      1094 KGLQSLPTHDPSLPQRYSEDPTVPLPS-----ETDGYVAPLTCSPQPEYVNOQDVPRQ 1146
Db      1112 RRPVAPHVQEDSTQYASDPYFAPERSPRGELDEGIMTMRDKPKQEYLNPEV--- 1167
Qy      1147 PPSPREGSLPAARPAATLERAKTLPSPGNVGVKDYFAFGAVENPEYLTPOGGAAPQH 1206
Db      1168 -----ENPFVSR-----KNGDLQ-----ALDPEYHNASNG----- 1194
Qy      1207 PPPA-----FSPAFDNIYYVDDQDPPERGA-- 1230
Db      1195 PPKADEYVNEPLYLNTFANTLGKAEYLNKNIILSMPEKAKKAFDNPDIWNHSLPPRSTLQ 1254
Qy      1231 PSTFKGRTPT-----AENPEYL 1247
Db      1255 HPDYLOEYSTKYFYKONGRIRPIVAVENPEYL 1285

```

```

Rp      SEQUENCE FROM N.A.
Rc      TISSUE=Heart;
Rc      MEDLINE=988221155; PubMed=9553078;
Ra      Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,
Ra      Marchionni M.A., Kelly R.A.;
Rt      "Neuregulins promote survival and growth of cardiac myocytes.
Rt      Persistence of ErbB2 and ErbB4 expression in neonatal and adult
Rt      ventricular myocytes."
Rl      J. Biol. Chem. 273:10261-10269(1998).
Rn      [2]
Rp      SEQUENCE OF 848-901 FROM N.A.
Rc      TISSUE=Sciatic nerve;
Rc      MEDLINE=9122560; PubMed=2025425;
Ra      Lai C., Lemke G.;
Rt      "An extended family of protein-tyrosine kinase genes differentially
Rt      expressed in the vertebrate nervous system."
Rl      Neuron 6:691-704(1991).
Rn      [3]
Rp      SEQUENCE OF 1031-1198 FROM N.A.
Rc      STRAIN=Sprague-Dawley; Tissue=Spinal cord;
Rc      MEDLINE=97184212; PubMed=9030624;
Ra      Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
Rt      "Expression of neuregulins and their putative receptors, ErbB2 and
Rt      ErbB3, is induced during Wallerian degeneration."
Rl      J. Neurosci. 17:1642-1659(1997).
Rn      [4]
Rp      FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
Rp      2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BTACELLULIN AND
Rp      NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
Rp      NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
Cc      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
Cc      tyrosine phosphate.
Cc      -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
Cc      RECEPTORS (POTENTIAL).
Cc      -1- SUBCELLULAR LOCATION: Type I membrane protein.
Cc      -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
Cc      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
Cc      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
Cc      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
Cc      HEART.
Cc      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
Cc      RESIDUES (BY SIMILARITY).
Cc      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
Cc      -----
Cc      This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc      between the Swiss Institute of Bioinformatics and the EMBL outstation -
Cc      the European Bioinformatics Institute. There are no restrictions on its
Cc      use by non-profit institutions as long as its content is in no way
Cc      modified and this statement is not removed. Usage by and for commercial
Cc      entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
Cc      or send an email to license@ebi.ac.uk).
Cc      -----
Cc      EMBL, AF041838; AAD08899.1;
Cc      EMBL, U52531; AAC53051.1;
Cc      HSPF, P11362; IFK.
Cc      InterPro: IPR000494; EGFR_L domain.
Cc      InterPro: IPR000719; Euk_DKINase.
Cc      InterPro: IPR002174; Furin-like.
Cc      InterPro: IPR001245; Tyr_KINase.
Cc      InterPro: IPR004019; YLP_motif.
Cc      Pfam: PF00757; Furin-like; 1.
Cc      Pfam: PF00069; PKINase; 1.
Cc      Pfam: PF01030; Recep_L_domain; 2.
Cc      Pfam: PF02757; YLP; 2.
Cc      PRINTS: PR00109; TYRKINASE.
Cc      ProDom: PD000001; Euk_DKINase; 1.
Cc      SMART, SM00261; FU; 4.
Cc      SMART, SM00219; TYRKC; 1.
Cc      PROSITE, PS00107; PROTEIN KINASE ATP; 1.
Cc      PROSITE, PS00011; PROTEIN KINASE DOM; 1.
Cc      PROSITE, PS00109; PROTEIN KINASE TYR; 1.
Cc      Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
Cc      Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
Cc      SIGNAL
Cc      1
Cc      25
Cc      POTENTIAL.

```

FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ENRB-4.  
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 652 675 POTENTIAL.  
 FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 166 334 CYS-RICH.  
 FT DOMAIN 496 633 PROTEIN KINASE.  
 FT DOMAIN 718 985 ATP (BY SIMILARITY).  
 FT NP\_BIND 724 732 ATP (BY SIMILARITY).  
 FT BINDING 751 751 BY SIMILARITY.  
 FT ACT\_SITE 843 843 BY SIMILARITY.  
 FT DISULFID 189 197 BY SIMILARITY.  
 FT DISULFID 193 205 BY SIMILARITY.  
 FT DISULFID 213 221 BY SIMILARITY.  
 FT DISULFID 217 225 BY SIMILARITY.  
 FT DISULFID 230 238 BY SIMILARITY.  
 FT DISULFID 234 246 BY SIMILARITY.  
 FT DISULFID 249 258 BY SIMILARITY.  
 FT DISULFID 262 289 BY SIMILARITY.  
 FT DISULFID 293 304 BY SIMILARITY.  
 FT DISULFID 308 323 BY SIMILARITY.  
 FT DISULFID 326 330 BY SIMILARITY.  
 FT DISULFID 503 512 BY SIMILARITY.  
 FT DISULFID 507 520 BY SIMILARITY.  
 FT DISULFID 523 532 BY SIMILARITY.  
 FT DISULFID 536 552 BY SIMILARITY.  
 FT DISULFID 555 569 BY SIMILARITY.  
 FT DISULFID 559 577 BY SIMILARITY.  
 FT DISULFID 580 589 BY SIMILARITY.  
 FT DISULFID 593 614 BY SIMILARITY.  
 FT DISULFID 617 625 BY SIMILARITY.  
 FT DISULFID 621 633 BY SIMILARITY.  
 FT MOD\_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 253 253 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 410 410 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 473 473 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 548 548 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CONFLICT 1062 1062 S -> N (IN REF. 3).  
 FT CONFLICT 1080 1080 PRT -> SYR (IN REF. 3).  
 SQ SEQUENCE 1308 AA; 146957 MW; D944B80996A08B41 CRC64;

Query Match 42.4%; Score 2889; DB 1; Length 1308;  
 Best Local Similarity 44.2%; Pred. No. 2,66-148;  
 Matches 598; Conservative 189; Mismatches 353; Indels 172; Gaps 28;

1 MELA-ALCRWGLT--ALLPPGAASVTCTGTDMKRLPLASPEHLDMLRHLYOGCCOVQ 57  
 1 MKLATGLMWSGLVARTVQPSASQVCACTENKSLSDLEQYALAKKYENECVVM 60  
 58 GNELTYLPTNASLFTLODIOEVGYLIANOVROVPLQRLVRGTOLFEDNYALAV 117  
 61 GNLEBTSIEHNRDLSFLRSIREVGYVALNOFRYPLENLRILIRGTLYEGRYALAI 120  
 118 DNGDPLNNTTPTVTGASPGJRELQRLSTELTKGVLIQRPOLCYODTLMDKIDPHKN 177  
 121 LNYRKDNF-----GLOELGKLTTELINGVYVDQNKFLCYADTIHWQDIVANPW 171  
 178 QVALTLIDTNSRACHPCSPCKSGRCGESSDDQSLTRTVCAAGC-ARCKGLPTDCC 236  
 172 PSNNITVSTIGSSGGRCHKSGCTG-RCWGPTEHNCQTLIRTVCAEQDDGRCYGYVSDCC 230  
 237 HEOGAAGCTGPKHSOCLALPHNSGICELHCPALVTVNTDTPESMNPBEGRYFGASCV 296  
 231 HRCAGGCGSKPDGDCFCACMNPDSGACVYQCQTFYVNTPTTFLEHNNAKYIYGAFCV 290

QY 297 TACPVNLTSDVSGCTLVCPDLHNOEYTAEDGTORCEKSKPCARVCVGLQMOYIKANSKF 356  
 DB 291 KCPHNV-VDSSCVACACSSKREV-EENIGIMCKRCTDIOCRACOGITGSLMSAQTV 348  
 QY 357 IGITELFPACCKIIFGLAFLPESFDGDPASNTAPLOEOLVPELTLEITGYIYSAMP 416  
 DB 349 DSNWIDFINTCKINGNLIFLVGIGHDPYNAIDADPEKXNFRTRYREITGFPLNTQTPW 408  
 417 DSLPDLVFOQLYIRRIILHNGAVSLTLQGLISWLGSLRELSGLLILHNTHLCF 476  
 409 PNMTDFSVSLVTGGVLYSGSLILKQOGLTSLOPSLEKISGNIIYITNSLVCY 468  
 477 VHTVPMQDLFRPHQALHTANRDEDECVBGLACHOLCARHCWGFPTQCVNCSQFLR 536  
 469 YHINMTLTFSTVQRLVIDNRRACNTAEGMVCNHLGNDGCGGPGPOCLSCRRFSR 528  
 537 GQCEVEBEVYQGLPREVYANRHLPGHPCQF-QNCSYTCFGEADQCVACAYKDPF 595  
 529 GKICIESCNLYDGFREFENGSLICVEDSQCKMEDGLLTHGHPDCTKCSHFKDGN 588  
 596 CVARCPGVRKPDLSYMPIWKEPDEEGACQPCPINCTHSCVDLDDKGC-----PA 644  
 589 CVERCPDVLQANF--IFKADQDRCHPCHPCTQCGNGPTSHDCIYPMWGTSLPQ 646  
 645 EORASPLTIVSAV-VGLLVTVLGVVPGILIFNNFTVSEFWLRPKYSASH-----LE 696  
 647 HAR-TPL--IAAGVIGLFLIVLALFAVYVRK-----SIRKKALRRFLETIVE 696  
 697 PLTBGAMPNOAORILKETELRKVKVLSGAFGVYKGIWIPGENVKIPVAIKVLREN 756  
 697 PLTBGAMPNOAORILKETELRKVKVLSGAFGVYKGIWIPGENVKIPVAIKVLREN 756  
 757 TSPKANKEILDEAVYMAVGVSPYVSRLLIGLCTSVOLVTLQMPYCLLDHVENRGLG 816  
 757 TSPKANKEILDEAVYMAVGVSPYVSRLLIGLCTSVOLVTLQMPYCLLDHVENRGLG 816  
 757 TSPKANKEILDEAVYMAVGVSPYVSRLLIGLCTSVOLVTLQMPYCLLDHVENRGLG 816  
 757 TSPKANKEILDEAVYMAVGVSPYVSRLLIGLCTSVOLVTLQMPYCLLDHVENRGLG 816  
 817 SODLLNMCQIAKAGSVLEDEYRIYARDLAANNVLYKSPNHNKIDPGLARLIDDETEYH 876  
 817 SODLLNMCQIAKAGSVLEDEYRIYARDLAANNVLYKSPNHNKIDPGLARLIDDETEYH 876  
 817 SODLLNMCQIAKAGSVLEDEYRIYARDLAANNVLYKSPNHNKIDPGLARLIDDETEYH 876  
 817 SODLLNMCQIAKAGSVLEDEYRIYARDLAANNVLYKSPNHNKIDPGLARLIDDETEYH 876  
 877 ADGGKVPILKMMALBESILRRPFTHSDVSVYGVYVWELMTFPAKPYDGI PAEIPDLLEK 936  
 877 ADGGKVPILKMMALBESILRRPFTHSDVSVYGVYVWELMTFPAKPYDGI PAEIPDLLEK 936  
 877 ADGGKVPILKMMALBESILRRPFTHSDVSVYGVYVWELMTFPAKPYDGI PAEIPDLLEK 936  
 877 ADGGKVPILKMMALBESILRRPFTHSDVSVYGVYVWELMTFPAKPYDGI PAEIPDLLEK 936  
 937 ERLPQPPICITIDVYIMVKCMWIDSECPREELVSEFSRMAKDDQRFVIONED-LGPA 995  
 937 ERLPQPPICITIDVYIMVKCMWIDSECPREELVSEFSRMAKDDQRFVIONED-LGPA 995  
 937 ERLPQPPICITIDVYIMVKCMWIDSECPREELVSEFSRMAKDDQRFVIONED-LGPA 995  
 937 ERLPQPPICITIDVYIMVKCMWIDSECPREELVSEFSRMAKDDQRFVIONED-LGPA 995  
 996 SPLDSTFVRSLEDDMDGLVDAEYLVPOQGFPCPD----- 1033  
 997 SBNDSKFPONLDEEDLEDMDAEYLVV-QAFNIPPIIYSTRHIDSRSEIGHSPRA 1055  
 1034 -APGAGVYHHRHRSSTRSGGGLTLGLEPSEEEAPSPPLASGAGSDVDGLMGCA 1092  
 1056 YTPMSGQCFVYODGFAITDQG--MPMEYATITSTIPAPVA--QGAFAEMDDSCNGT 1110  
 1093 AKGLDLSLPHDSPLORSEDPVPLPS-----ETDGYVFLTCSPOPEVNOQDVAP 1145  
 1111 LKPVVPHVQEDSSQRIYADPTVAPAEKNPAAELDEBQYVPMHDKRQVELNVE--- 1167  
 1146 QPSPREGRPLAPARAGATLERAKTSLSPKNGVXKDVAFGAGAVENPEYLTQGAAPOP 1205  
 1168 -----ENFVSR--KNNDLQ-----ALNPEVHSSASG----- 1194  
 1206 HPPRA-----FSPAPNLVYWDQDPPERGA- 1230  
 1195 -PPKAEDEVYNEPLVINTTALGNAEYKNSLSVPEKAKAFNDPVMHNSLPPRSTL 1253  
 1231 -PPSTFKGTPT-----AENPEYL 1247  
 1254 QHPDYLAQEVSTKRYKQNGRIRPIVAENEYLL 1285

RESULT 8  
XMRK\_XIPMA STANDARD; PRT; 1167 AA.

AC P1388;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).  
GN XMRK OR TU.  
OS Xiphophorus maculatus (Southern platyfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
OX NCBI\_Taxid=8083;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90015140; PubMed=2797166;  
RA Witbrodt J., Adam D., Malleschek B., Mauelel W., Raulf F.,  
RA Telling A., Robertson S.M., Schartl M.;  
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-  
inducing Tu locus in Xiphophorus."  
RL Nature 341:415-421(1989).  
RN [2]  
RP REVISION TO 515.  
RA Schartl M.;  
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
CC - FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.  
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
tyrosine phosphate.  
CC - SUBCELLULAR LOCATION: Type I membrane protein.  
CC - DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.  
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
at the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL: X16891; CAA34770.2; -  
DR PIR: S06142; S06142.  
DR HSSP: P11362; 1FGK.  
DR InterPro: IPR000494; EGFR\_L\_domain.  
DR InterPro: IPR000719; Euk\_pkinase.  
DR InterPro: IPR002174; Furin-like.  
DR InterPro: IPR002290; Ser Thr\_pkinase.  
DR InterPro: IPR001245; Tyr\_pkinase.  
DR Pfam: PF007657; Furin-like.1.  
DR Pfam: PF01030; Recep\_Ldomain.2.  
DR PRINTS: PR00109; TYRKINASE.  
DR ProDom: PD000001; Euk\_pkinase.1.  
DR SMART: SM00261; FU; 5.  
DR SMART: SM00220; S\_TKC.1.  
DR SMART: SM00219; TYTKC.1.  
DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP.1.  
DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR.1.  
DR PROSITE: PSS0011; PROTEIN\_KINASE\_DOM.1.  
KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;  
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.  
FT SIGNAL 1 25  
FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE  
KINASE  
EXTRACELLULAR (POTENTIAL).  
FT DOMAIN 26 642  
TRANSMEM 643 665  
FT DOMAIN 666 1167 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 710 977 PROTEIN KINASE.  
FT NP BIND 716 724 ATP (BY SIMILARITY).  
FT BINDING 743 743 ATP (BY SIMILARITY).  
FT ACT\_SITE 835 835 BY SIMILARITY.

FT DISULFID 195 204 BY SIMILARITY.  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 220 228 BY SIMILARITY.  
FT DISULFID 224 236 BY SIMILARITY.  
FT DISULFID 237 245 BY SIMILARITY.  
FT DISULFID 241 253 BY SIMILARITY.  
FT DISULFID 256 265 BY SIMILARITY.  
FT DISULFID 269 296 BY SIMILARITY.  
FT DISULFID 300 311 BY SIMILARITY.  
FT DISULFID 315 330 BY SIMILARITY.  
FT DISULFID 333 337 BY SIMILARITY.  
FT DISULFID 504 513 BY SIMILARITY.  
FT DISULFID 508 521 BY SIMILARITY.  
FT DISULFID 524 533 BY SIMILARITY.  
FT DISULFID 537 553 BY SIMILARITY.  
FT DISULFID 556 569 BY SIMILARITY.  
FT DISULFID 560 577 BY SIMILARITY.  
FT DISULFID 593 615 BY SIMILARITY.  
FT DISULFID 618 626 BY SIMILARITY.  
FT DISULFID 622 634 BY SIMILARITY.  
FT DISULFID 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

Query Match 38.4%; Score 2614.5; DB 1; Length 1167;  
Best Local Similarity 44.2%; Pred. No. 1.4e-133;  
Matches 562; Conservative 168; Mismatches 394; Indels 147; Gaps 30;

4 AALCWGULLALPFGAST---QVCTGTMKRLRSPSPETHLDMRLHYOGCVOGQGN 59  
8 AALLD--LLVLVISIRCCSTDPDRVCQGSITQNM--LDNHVYKMKKMSGCVAVLEN 62  
60 LETLPTPNASISPLDIOEVQGYVLIAHNOVRQVPLRLRIVRGTOLEFEDNYALAVLDN 119  
63 LEITYGENODLSFIQSQEVGYVLIAHNEVSTPLVNLRLIRQNLIEGNFTLLVMSN 122  
120 GDLPLNTPVTGASGCGIRELOLSLTILKGVLIQNPOLCYODTILMKDIFPKKNQL 179  
123 YOK-NPSSP--DVYVGKQLOLSLTILKGVVSHNPPLCNVETINMMDIVKTSNP 179  
180 ALLIDTNRSPRCHPSCPMCKGSRGWSSESDCGSLTRVAGGC-ARCKGPLPDCCH 238  
180 TMLILPHAFERQCKCHGCVNNGSCMAFGPHGCKFTLLCAEQCNRCRGKPIDCNE 239  
229 OCACCTGPKHSDCLAHFNHSGICELHCPALVYNTDTESMNPBGRYTFGASCYTA 298  
239 OCACCTGPKHSDCLAHFNHSGICELHCPALVYNTDTESMNPBGRYTFGASCYTA 298  
240 HCAGGCTGPRATDCLACRDFNDGCTCKPTCYDVSQVVDNPNIKYTFGAACVKE 299  
299 CPYNYLTDVSGCTLVCLHNGEVAEDGTQRCCKSCPCARVCYGLQMOYIKANSKFG 358  
300 CPSNVTVE-GACVNSGSGAMLEVD-ENGKSKCKDGVCPKVGIGIST-SVTIVN 356  
359 ITEL-EFAGCKKIFGSLAFDPESPDGDASTAPLQPOLQVFEFTLEBITGYLYISAMP 417  
357 STNHSFNSCKTINDIILNNSFEGDHYKIGTMDPHLMLLTVEKITGYLYIMMPE 416  
418 SLPLDSVQNLQVITGRILHNGAYS-LTLOGIGISLWELRLSRELGSGLALIHNTLCLF 476  
417 NMTSLSVQNLLEIRGRITTFGRGFSFVAVVQVRLQWELRLSRELGSAGANVLKNTLQRY 476  
477 VAVVMDQLFRNPHQALHTANRPEDECVGEGALCHQCARGHGCGPPTOCVNSCPLR 536  
477 ANTIMWRLLFNSDEDSIYDART-----ENQTCNNSGSDCGPPTOCVNSGLAHDR 529  
537 GQECVEECRVLQGLPREYVNAHCLCPHPCOPQNGSVTCGPEADQCVACAHYKDPFC 596

Db 530 GAGCAGVASCNLLQEPREAVDGRVCHQCELVQTSLTCTGCGPANCSAFAFDGPOC 569  
 Qy 597 VAPCPGKPDLSYMPIMKPEDEGACPCPINTGSCVLDKGCAPGASPLTGISVS 656  
 Db 590 IPRCPHILDBDGLT-LMKYADKMGCGCPCHQNTGCGSPGLSGCGD-IVSHSLAVG 647  
 Qy 657 AVNGILLVVLGVVFGIL-----IENFTYSFMRVVKVASHLEPITPGAMPNOAQR 711  
 Db 648 LVSGLLITVALLIIVLLRRRIKRRRTICLQEKEL-----VEPLTSGQAPNOAFLE 703  
 Qy 712 ILKETELKVKVLSGAGFVYKGIWIPDEGNVKAIPAIKYLENTSPKANKELDEAYV 771  
 Db 704 ILKETEKQKRVLSGAGFVYKGIWIPDEGNVKAIPAIKYLENTSPKANKELDEAYV 763  
 Qy 772 MAGVSPVSRLLGLCTSTVQVLTQMLPFGCLLDHRENGRGLSGDILNMCQIAKGM 831  
 Db 764 MASVDHPVCRLLICITLSAVQVLTQMLPFGCLLDHRENGRGLSGDILNMCQIAKGM 823  
 Qy 832 SYLEDVRLVHRDLAARVNLVKSPPHVKITDPGLARLDIDETEHADGKVPKIMMALES 891  
 Db 824 NYLEERHLVHRDLAARVNLVKSPPHVKITDPGLARLDIDETEHADGKVPKIMMALES 883  
 Qy 892 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPABIPDLEKGERLPPOPICTIDVYM 951  
 Db 884 ILQWTFHQSDVMSYGVTVWELMTFGSKPYDGIIPAKIASYLENGERLPPOPICTIEVYM 943  
 Qy 952 IMVCKMTDSCRPRFELVSEFSRMAPDQRFVIONEDLGAPSLDSTFFYSLEDDDD 1011  
 Db 944 ILLCKMTDSCRPRFELVSEFSRMAPDQRFVIONEDLGAPSLDSTFFYSLEDDDD 1000  
 Qy 1012 MGDLVDAEYLVPOOGFPCPDPAAGAGMVRHRSSTRSGGDLTLGLEPSEEARPS 1071  
 Db 1001 -DVVDADEYLLPYKRI-----NRQGS-----E 1021  
 Qy 1072 PLASEGAGSDVPFGDGLMGAAKGLQSLPTHDSPRLQRYSDPPIV-PLPSTGCVYAPLT 1130  
 Db 1022 PCIPPTGH-----PVRENSITLRNISDPQNALMKDLQGH----- 1056  
 Qy 1131 CSPOEYVNPQDVPRQP-----PSPRE-----GPUP-AARPAAGATLERAKTLPSPKNG 1177  
 Db 1057 -----EYVNPQSESSLSLSDIYNPNVEDLTDGMPVSLSSQAEITNPSREYLTNQN 1111  
 Qy 1178 VKQVFAFGAVENDEYLTPOGGAAPQHPHPPASPAFDNLTYWDQDPPEKAPPSFFKG 1237  
 Db 1112 L-PLVSSGSMDDPDY---QAG-----YQAP-----LPQTGALTGNGMF 1146  
 Qy 1238 TPTAENPEYLG 1248  
 Db 1147 LPAENLEYL 1157

RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90311312; PubMed=2164210;  
 RA Tolman G.D., Whittney G.S., Neubauer M.G., Green J.M., McDonald V.L.,  
 RT Todaro G.J., Shoyab M.;  
 RT "Molecular cloning and expression of an additional epidermal growth  
 factor receptor-related gene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).  
 RN [3]  
 RP SEQUENCE FROM N.A. (SHORT FORM).  
 RC TISSUE=Placenta;  
 RX MEDLINE=93282822; PubMed=7685162;  
 RA Katoh M., Yazaki Y., Sugimura T., Terada M.;  
 RT "c-erbB3 gene encodes secreted as well as transmembrane receptor  
 tyrosine kinase";  
 RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).  
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTRK.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND  
 CC SECRETED (SHORT FORM).  
 CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A  
 CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.  
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
 CC AND PROMOTES ITS ASSOCIATION WITH THE P65 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).  
 CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M29366; AAA35790.1; -  
 DR EMBL; M34309; AAA35979.1; -  
 DR EMBL; S61953; AAB26935.1; -  
 DR PIR; A36223; A36223.  
 DR HSSP; P11362; 1FGK.  
 DR Genew; HGNC:3431; ERBB3.  
 DR MIM; 190151; -  
 DR InterPro; IPR000494; EGF\_L\_domain.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_Pkinase.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR ProDom; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; TYRK; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; FALSE NEG.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; FALSE NEG.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Alternative splicing.  
 KM SIGNAL 1 19  
 FT CHAIN 20 1342 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-3.  
 FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 644 664 POTENTIAL.  
 FT DOMAIN 665 1342 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 709 966 PROTEIN KINASE.  
 FT NP\_BIND 715 723 ATP (BY SIMILARITY).



```

FT BINDING 742 742 ATP (BY SIMILARITY).
FT ACT SITE 834 834
FT DISULFID 186 194 BY SIMILARITY.
FT DISULFID 190 202 BY SIMILARITY.
FT DISULFID 210 218 BY SIMILARITY.
FT DISULFID 214 226 BY SIMILARITY.
FT DISULFID 227 235 BY SIMILARITY.
FT DISULFID 231 243 BY SIMILARITY.
FT DISULFID 246 255 BY SIMILARITY.
FT DISULFID 259 266 BY SIMILARITY.
FT DISULFID 290 301 BY SIMILARITY.
FT DISULFID 305 320 BY SIMILARITY.
FT DISULFID 323 327 BY SIMILARITY.
FT DISULFID 500 509 BY SIMILARITY.
FT DISULFID 504 517 BY SIMILARITY.
FT DISULFID 520 529 BY SIMILARITY.
FT DISULFID 533 549 BY SIMILARITY.
FT DISULFID 552 565 BY SIMILARITY.
FT DISULFID 576 585 BY SIMILARITY.
FT DISULFID 589 610 BY SIMILARITY.
FT DISULFID 613 621 BY SIMILARITY.
FT DISULFID 617 629 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA: 148097 MW: 7201E7F66CA374BD CRC64;

Query Match 35.1%; Score 2389.5; DB 1; Length 1342;
Best Local Similarity 40.0%; Pred. No. 2.2e-121;
Matches 527; Conservative 195; Mismatches 453; Indels 141; Gaps 35;

QY 10 GLLALLPAGA--STQVCTGTDKLRLPASPTHLMLRHLVQGCQVVGQNLDTLPT 67
DB 11 GLLFSLARGESEVNSQAVCEPTLNGLSVTGDAENQYOTLYKLYERCEVVMGNLEIVLTGH 70
QY 68 NASISELQDIQEVQGVYLAHNOVROVRLRIVRGTOLEFEDNYALVALDNDPLNNTT 127
DB 71 NADISFLQWIREVTVLVANNEFSTLPLRLRVKRGTOVYDGFALFVN----LNYNT 125
QY 128 PVTGASPGRLRELRLRLTEILKGLVLIQNPOLCYODTILMNDIFKNNQALTLIDTN 187
DB 126 ----NSSHALRQRLTQLTEILSSGVYIEKNDKLCMDTIDMRDIYDRD--AEIVKD 178
QY 188 RSPACHPCSMCKSRKMGWGESSEDCSLTRTVCAAGC-AACKGLPLDCHCEQAACGTS 246
DB 179 NGRSPCPCHVEYCKG-KMGWGESSEDCSLTRTVCAAGC-AACKGLPLDCHCEQAACGTS 237
QY 247 PKISDCLACLHFNHSGICELHCPALVTYNTDTEFSPMBNPEGRYTFGASCYTAAPYVLTST 306
DB 238 PQTDTCPACHRNHNDSGACVPRCPQLVYNNKLTFLBENPHTKYQGVGVCAVSPHNHV-V 296
QY 307 DVSGCTVCPULHNOEYVAEDGTQCEKCSKPCARVVCGLMQYIKANSKR--IGITELF- 363
DB 297 DQTSQVACPEPDKKEVD-KGKMKCECGGLCPKACGSGT-----SSSRQTYDVSSNDG 350
QY 364 FAGCKKIFGSLAFPLPESFSDGPASNTAPLQPEQLOVETLEITGYIYIAMPDSDPLDS 423
DB 351 FVNCSTKLGNLDFILTGNDPWHKIKALDPEKLNVRTRVREITGYINISWEPHNHNS 410

```

```

QY 424 VFQNTQVIRGRILHNGAAS-LTLQGLIGISMLGLRLRELSGLALIHNTLCEVHTVPM 482
DB 411 VFSNLTITGSGSLVNRGSLIMKLNANTSLGFRSLKELISNRITISANRQICNHSLNW 470
QY 483 DQLEFNPHQALHTA-NRPEDECVGEGLACHQLCARGHCWGPPTQCNCSQFLRGQCV 541
DB 471 TVVLGRPTREERDIDKHNRRPRDCVAEGRYCDBPLCSSGCGMGPGQCCISCRNYSRGVGV 530
QY 542 EECRVLQGLIPREYVNAARCLPCHPECPQNSVTCFGEADQCVACAHYKDPFCVAPCR 601
DB 531 THCNFLNGEPREFAHEABCFSCHEPCQMEGATNGSGSDTCAQACAFRDPCHVSSCP 590
QY 602 SGVCKPDLSPMPKPEDEGAQCPICNTHSCVDLDDKCPAEQRA----SPLTISVA 657
DB 591 HGVLG-AKGPYKXPDVQNECRPCHEKCTQCKRPELDCLGTLVLIGKTHLMALTY 648
QY 658 VVGILLV-VLGVV-----GILFNFTVSPFLRVKXSASLLEPLTSGAMPNOQMFI 712
DB 649 IAGLVVIFEMVLGSLVLYMRGRRIQKRAMRYLE---RGESIEPLDPS-EKANKVLAHI 703
QY 713 LKETELRKVVLGSGAFGVYKGIWIPGENYKIPVAIKVIRENTSPRANKELDEAYVM 772
DB 704 FKETELRKVLGSGVFGVYKGIWIPGESIKIVCIKVIDKSGRQSFQAVTDMALAI 763
QY 773 AGVSPVYSRLIGLCTSTVOLVQLMPYGLLDHVENRGRFLSGQDLNMCQIAKMS 832
DB 764 GSLDHAHIVRLIGLPGSSSLQVLTQYLPGLSGSLDHHVROHKGALGQLLMGVOIAKGVY 823
QY 833 YLEDRIYVRDLAENVLYKSPNHKINDFGARLLDDETEYHNDGGVPRKMMALESI 892
DB 824 YLEEHQWHRNLAAKNVLLKSPSYQVADFGADLLPDDKOLLYSEAKTPIKMMALESI 883
QY 893 LRRPFTQSDVNSYGVTVWELMTFGAKPYDGIIPAREIDLEKGERLPDPICTIDVMI 952
DB 884 HRGKTHQSDVNSYGVTVWELMTFGAEPAVGLRLAEVPLLEKGRLAQPCITIDVMI 943
QY 953 MYKCMWIDSECPRERELVSFEMARPDQRFVYQNEDLGA--SPLDSTFYSLLED 1009
DB 944 MYKCMWIDSECPRERELVSFEMARPDQRFVYQNEDLGA--SPLDSTFYSLLED 1002
QY 1010 DDMGDLVDAEYLVQCGFPCDPAPAGAGWNNHRSSTSGSGDILLGLEP-SEEA 1068
DB 1003 VLEPELDLDDLEAED-----NATTLTSLSLPSTGLNRRPG 1043
QY 1069 PRSPPLAPSEGASDVFDGLGKAAGLQSLPTH-PSFLQRYSEDPTVPLP-----SE 1121
DB 1044 SOLSLSPSSGY-MPMNQNLGESCQBSAVSGSSERCPVSLH-----PMRGCIASES 1096
QY 1122 TDGVYA-----PLCSPPQE-----YVNPQVVRPQPSRREGP----- 1154
DB 1097 SEGHTVGSFAELQEKVSMKRSRBSRSPRPGDSAYHSORHSLTLPTPLSPGLEEDV 1156
QY 1155 ----LPAARPAATLERAKTISP-GKNGVY-----KVFAGAVENBEYVLPQGAAP 1203
DB 1157 NGYVMDPTLTKGTSPSREGLTSSVGLASVLTGEEDED-----EEVEYNNRRRHSP 1208
QY 1204 QHPPPAFSPAFDNLVYMD-----QDPREKRGAPSTFTGTPAENPEYL 1247
DB 1209 -PHPPPSLEELGYEYMDVSGDLSASLSTGSCPLHPVLPMTAGTTPDEDEYEM 1263

RESULT 10
ERR3 RAT STANDARD; PRT; 1339 AA.
AC 062759; 062955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; Tissue=Liver;  
 RX MEDLINE=96096535; PubMed=8522190;  
 RA Heliyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;  
 RT "Cloning of the rat ErbB3 cDNA and characterization of the  
 RL recombinant protein.";  
 RL Gene 165:279-284(1995).  
 RN [2]  
 RP REVISIONS TO 85: 513 And 565.  
 RA Heliyer N.J., Koland J.G.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 922-1097 FROM N.A.  
 RC STRAIN=Sprague-Dawley; Tissue=Sciatic nerve;  
 RX MEDLINE=97184212; PubMed=9030624;  
 RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;  
 RT "Expression of neuroligin and their putative receptors, ErbB2 and  
 RL ErbB3, is induced during Wallerian degeneration.";  
 RL J. Neurosci. 17:1642-1659(1997).  
 CC -1 FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.  
 CC -1 CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1 SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -1 SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1 DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
 CC -1 PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE.  
 CC -1 SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

FT DISULFID 190 202 BY SIMILARITY.  
 FT DISULFID 210 218 BY SIMILARITY.  
 FT DISULFID 214 226 BY SIMILARITY.  
 FT DISULFID 227 235 BY SIMILARITY.  
 FT DISULFID 231 243 BY SIMILARITY.  
 FT DISULFID 246 255 BY SIMILARITY.  
 FT DISULFID 259 286 BY SIMILARITY.  
 FT DISULFID 290 301 BY SIMILARITY.  
 FT DISULFID 305 320 BY SIMILARITY.  
 FT DISULFID 323 327 BY SIMILARITY.  
 FT DISULFID 500 509 BY SIMILARITY.  
 FT DISULFID 504 517 BY SIMILARITY.  
 FT DISULFID 520 529 BY SIMILARITY.  
 FT DISULFID 533 549 BY SIMILARITY.  
 FT DISULFID 556 573 BY SIMILARITY.  
 FT DISULFID 576 585 BY SIMILARITY.  
 FT DISULFID 589 610 BY SIMILARITY.  
 FT DISULFID 613 621 BY SIMILARITY.  
 FT DISULFID 617 629 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CONFLICT 1028 1028 L->P (IN REF. 3).  
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402B5FD1E CRC64;  
 Query Match 34.0%; Score 2315.5; DB 1; Length 1339;  
 Best Local Similarity 40.4%; Pred. No. 2.1e-117;  
 Matches 520; Conservative 171; Mismatches 431; Indels 165; Gaps 36;  
 QY 3 LALCRMGILLALLPPGAA---STQVCTGTDMKRLRPASPEHLDMLHLYGCGQVQVGN 59  
 DB 7 LQVLC---FLSLARGSEMGNSQAVCPDTLGLSLVTDADQVQVTLKLYKECVVGN 62  
 QY 60 LELTLPYASLFIQDIOEVGYLILANQVRQVPLRLIRVGTOLFEEDVLAVALVDN 119  
 DB 63 LEIVLGNHADSFLQWIREVGYLVANRESVPLPLRLRVRGTOYDQKFAIFVM-- 120  
 QY 120 GDPLNNTPTVPGASPGELRELQRLRELTILKGVTLIQRPOLCYDPTILMKDIFKKNOL 179  
 DB 121 ---LNVNT---NSSHALRQLKFTLTETLSGVYIEKXDKLCHMDTIDMRDIVVR-- 170  
 QY 180 ALLTIDTNSRACHPCSPMCKSGRCMGSESSDQSLTRTVACGGC-ARCKGPLPTDCHE 238  
 DB 171 GAETIVKNGANCPCHEVCKG-RCKGPPDDCQILITKTCAPQCNCRGFGNPNQCCHD 229  
 QY 239 QCAAGCTGPKSHDCLALFHNSGICELHCPALVTYNTDFESMNPREGRYFGASCYTA 298  
 DB 230 ECAGGSGQPDITDFCAFRFRFNSGACVPCPEPLVYNNKTLFQLEPNPHTKYGGVCVAS 289  
 QY 299 CPNNYIISTVSGCTIVLCPILHNOEYVAEDSTQCEKSGKRCQAVCYGL--GMQYIANSKF 356  
 DB 290 CPENFY-VQOTFCVRACPDKMEVD-KHLLKMCCEPGGCLCPACGTGSGASYQVDSN 347  
 QY 357 IGTLEFPAQCKKIFGSLAFLESPFDGPNASTAPLOEQLOVFTELEITYLYISAWP 416  
 DB 348 ID---GFVNCRTKIIGLNDLFLTGLVDMHKI PALDEPKLVFVREITGYLNIQSWP 403  
 QY 417 DSLPDLISVQONQVIRGRILNHGANS-LTLOGLISWTLGRLRELGSLALIHNTLHC 475  
 DB 404 PHHNSVSFNSLTTIGRSLYNGSFLIMKPLNVTSGFRSLKTSISARVYISANQOQC 463  
 QY 476 FVATVPMQDLFRNPHQALHTA-NRPEDECVGEGLACHOLCARCGCMGPGPTQVCNCSQF 534  
 DB 464 YHSLNMTTLTLGSPSERIDIKYDRPLSGCLAEKGVCPDLSSGGCGMGPGQCLSCNRY 523  
 QY 535 LRQQCEVEBCRYLQGLPREYVNAHCLPCHECQFONGSVYTCFGEADQCAVACAYKQDP 594

Db 524 SRGVCVTHNFIQGEREFVHQAQCSCHPECLPMEGISTCNGSSDACARCAHRCOP 583  
 QY 595 FCVACRPSGVKPLSTWPIKFPDEGACOPCPINCTHSC--VDLDDKCPAQRASPLT 652  
 Db 584 HCVNSCPHGLG--AKGPIKYDPAQNECRPHCHENTCGNGPELDDCQAVALNSKPH 641  
 QY 653 SIYSAVYGI-LLVVLGVVF---GILFNFTVSFWLRVPKYSASHLEPLTSGAMPNQ 707  
 Db 642 LVIAVYGLAVIILMILGSGFLYWRGRIRQKRAMRYLE---RGSISFLDSS-EKAKK 696  
 QY 708 AQMRILKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPANKIILD 767  
 Db 697 VLARIFKETELRKVKVLGSAFGTVYKGIWIPDGENVKIPVAIKVIRENTSPANKIILD 756  
 QY 768 EAVVMAGVSPVYSRLIGLCTSTVOLVQTMVYGCILDHVRENRRRLGSDPLNNCMOI 827  
 Db 757 HMLAVGSLDPAHIVRLGLCPGSSISLQVLYOYLPGLSLDHVKHRETLGQILLMNGVQI 816  
 QY 828 AKGMSYLEDVRLVHRDLAARNVLVKSPNHVKITDFGLARLLDIDETRYADGKVPDKM 887  
 Db 817 AKGMYLDEHSMSVHRDLARNWMLKSPSQVQVADDFVADLLPDDKQLLHSEAKTPIKMM 876  
 QY 888 ALSEILRRFTHOSDWSYGVWELMTFGAKPYDGPAPETIDLEKGRLEPQICIT 947  
 Db 877 ALSEIHFGKTHOSDWSYGVWELMTFGAEYPAGRLAEIPDLLEKGRLEAQPICIT 936  
 QY 948 DVTMIMVCMWIDSECRPRELVEFSRMAPDPORFVVIQNEIDLGPASPLDSTFRRSL 1007  
 Db 937 DVTMIMVCMWIDSECRPRELVEFSRMAPDPORFVVIQNEIDLGPASPLDSTFRRSL 993  
 QY 1008 EDDMDGLVDAEYLVPQGFPCPDPAFGAGVHRRHSSSTRSGGDLTLGEPSEF 1066  
 Db 994 TTREL-----QEALEPEL-----DLDDLLEAEEG 1019  
 QY 1067 -----BAPSPPLAPSEG-----AGSDVFDGLGMAKAGLQSL 1099  
 Db 1020 LATSLSGALSPLTGLTLRPGSGLSPSSGVMFMNOSLGEACLDASVAGSGEQSRPT 1079  
 QY 1100 PTHDPSPLQRYSEDFVPLPSEFDGVY---ADP-----TC-----SPQEP---YVN 1139  
 Db 1080 SLH-FIPGR-----PASESEGHVTSGEALOEKVSVCRSRSRSPRPGSDAAYS 1131  
 QY 1140 QPDVPPQPSPPREG-----LPARPAQATLBRATKLTSP-GKXGV-----CD 1181  
 Db 1132 QHSLSLTPVPLSPGLSEEDGNGYMPDTHLKGASSRECTLSVGLTEBEDD 1191  
 QY 1182 VFAFGAVENPEYLLTPQGAAPQPHPP 1208  
 Db 1192 -----EVEYVNRKRGRSP-PRPP 1209

RESULT 11  
 EGFR\_DROME STANDARD; PRT; 1426 AA.  
 AC P04412; O61601; Q9W2G0; P81868;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EG 2.7.1.112) (Egfr)  
 GN (Gurkin receptor) (Tropo receptor protein) (Drosophila relative of ESRB).  
 OS EGFR OR TOP OR C-ERBB OR DER OR CG10079.  
 OS Drosophila melanogaster (fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 CC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN  
 RP SEQUENCE FROM N.A. (ISOFORM TYPES I AND II).  
 RX MEDLINE=94350209; Pubmed=8070664;  
 RA Clifford R., Schupbach T.;  
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals  
 that several genetically defined classes of alleles cluster in

RT subdomains of the receptor protein."  
 RL Genetics 137:531-550(1994).  
 RN [2]  
 RP REVISIONS.  
 RA Clifford R., Schupbach T.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE=85124611; Pubmed=2982499;  
 RA Lively E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;  
 RT "The Drosophila EGF receptor gene homolog: conservation of both  
 RT hormone binding and kinase domains."  
 RL Cell 40:599-607(1985).  
 RN [4]  
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.  
 RC STRAIN=Oregon-R; Tissue=Embryo;  
 RX MEDLINE=87002474; Pubmed=3093080;  
 RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;  
 RT "Alternative 5' exons and tissue-specific expression of the  
 RT Drosophila EGF receptor homolog transcripts."  
 RL Cell 46:1091-1101(1986).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=99102120; Pubmed=9882502;  
 RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;  
 RT "Several levels of EGF receptor signaling during photoreceptor  
 RT specification in wild-type, Ellipse, and null mutant Drosophila."  
 RL Dev. Biol. 205:129-144(1999).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I).  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; Pubmed=10731132;  
 RA Adams M.D., Celinkner S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Aamodt C.P., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blake R.G., Champe O., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkuh C., Baldwin D.,  
 RA Bailey R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benca P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burris K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lascko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Mikhina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pachet J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington R.D.C., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spierker R., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Switzkes R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Xu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster."  
 RL Science 287:2185-2195(2000).  
 RN [7]  
 RP SEQUENCE OF 959-1078 FROM N.A.

[illegible]

```

Db      543 RWAFAIKPEPQKAWVWVENLADLCENKNGTICSDQCNEDGQWGWGTDGCLTKNFNNGTC 602
Qy      541 VEECRVLQGLPREYVNAARHCLPHREPCQPNQSVYTRGPAPDQCVACAHYKDPPEVVAAC 600
Db      603 IADQGISNAVK--FDNRCKIKIHPECR-----TCNGAGADHCQECVHRDQHCVSBC 654
Qy      601 P-----SGVK-----PDL 608
Db      655 PKKKYNDRGVCRECHATCDGCTGPKDTIGIAGCTTCNALINNDATVKKCLKDDKCPD- 713
Qy      609 SYMPWKE--PDEEGACQ-----CPI-----NCTH----- 632
Db      714 GY--FMEYVHPEQGSIKPLAGRAVCRKCHPLCELCTNYGHQVCSKTHYKREQCET 771
Qy      633 -----SC-----VLDLKG----- 641
Db      772 ECPADHYTDEEQECCFORHPECNCGCTGPGADCKSCNFKLFPANETGPNVASTMNCST 831
Qy      642 -CEAQR-----ASPLTSIVSAVGI--LVVVLGVVFGILLFNNFTVSEW 684
Db      832 KCPLEMHVNYQYTAIGPYCAASPRESSKITANLDVMMFIITGAVLVPTICLCVTVYI 891
Qy      685 LRPKYSASHL-----EPITSGAIPNQOMIILKETLKKYKUGSAFET 731
Db      892 CRQKQKAKKETVKMTALSCGDESEPLRPISNIGANLCKLTVDAELRKGVLWGAFGR 951
Qy      732 VYKGIWIPDGENYKIPVAIKVLRNTPSKANKEILDEAYVAVAGVSPYVRLIGLCTST 791
Db      952 VYKGVWVPEGENYKIPVAIKELKSTGASSEFLEHAYIMASEHVNLLKLLAYCMSQ 1011
Qy      792 VQVLQMPYGCILLDHYRENRKGLSGODLLNMCQIAKMSYLEDVLYHRDLAARVLY 851
Db      1012 IMLITQMPGLCLLDYVRNRRDKIGSKALNMSTQIAKMSYEEKRLVHRDLAARVLY 1071
Qy      852 KSPNHYKITDFGLARLDIDETEVHADGKVPKMMALSEILRRFTHOSDVSYGVTVW 911
Db      1072 QTSLSVKITDFGLAKLSSDSNRYKAAAGKMPKMLALBEIRKRVTSKSDVAHFGVTTW 1131
Qy      912 ELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPCTIDVYIMWKCMIDSECRPRELEV 971
Db      1132 ELTTFGQRPHEINIPAKIDPLIEVGLLEQPELCSIDYCTLLSCHMLDAAMPFKQULT 1191
Qy      972 SEFSMARDDPQRFVIVIONEDLG--PASPLDSTFYRSLIEDD--DWGDVDAEYLVPOQ 1026
Db      1192 TVFAEFARDDGRILALPGDKFTLPA-----YTSODEKDLIRKLAPTTDGEALAKPD 1244
Qy      1027 GFECPPDPAFGAGGVHRRHRSSTRSGGDLTLGLEPSEEAR-----RSLAPSEGAG 1080
Db      1245 DYLOPKAAPGPS-----HRTDCT-----DEMFKLNRVCMDPKNKSSGTG 1283
Qy      1081 SDVFDG---DLGMAKAGLQSLPTHDSPLQRYSEDPVTVLPSETQGYAAPLTGSPQPEY 1137
Db      1284 DDERDSAREVAGNENR-----LPLPVDEDYDLMF--TCQGGPNN 1321
Qy      1138 VNQDVPAPQPSREGEPLPAARPAATLEBAKTLSPGKGVNVDAVFAFGAVNENPEYL- 1195
Db      1322 NNNNN-----NPNQNNMAVVAAGYM-----DLGVPVSDNPEYLLN 1360
Qy      1196 --TPQGAAPQPH-----PPAPSP-AFDNLVYWD 1222
Db      1361 AQLTGVGESEIPTQITIGIPVWGGPTWEMVPMGPSEPTSSDHEYND 1408

```

```

OS      Avian leukosis virus.
OC      Viruses: Retroid viruses; Retroviridae; Alpharetrovirus.
OX      NCBI_Taxid=11864;
RN      (1)
RP      SEQUENCE FROM N.A.
RX      MEDLINE=85282822; PubMed=2988784;
RA      Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman F.M.,
RA      Crittenden L.B., Raines M.A., Kung H.-J.;
RT      "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT      processing and promoter insertion result in expression of an
RT      amino-terminated EGF receptor."
RL      Cell 41:719-726 (1985).
CC      -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC      IN CHICKS WHICH HAVE BEEN INFECTED WITH THE AVIAN LEUKOSIS VIRUS
CC      AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC      THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC      C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC      -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC      PROTEIN.
CC      -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL: M10066; AAA8763.1; ALT_INIT.
CC      PIR: A00643; TVCHLV.
CC      PIR: B00643; TVFVLV.
CC      HSSP: P1362; 1FGK.
CC      InterPro: IPR000719; Euk PKinase.
CC      InterPro: IPR001245; Tyr_Pkinase.
CC      Pfam: PF00069; PKinase; 1.
CC      PRINTS: PR00109; TYRKINASE.
CC      ProDom: PD000001; Euk PKinase; 1.
CC      SMART: SM00219; Tyrc; 1.
CC      PROSITE: PS00107; PROTEIN KINASE ATP; 1.
CC      PROSITE: PS00109; PROTEIN KINASE TYR; 1.
CC      PROSITE: PS00111; PROTEIN KINASE_DOW; 1.
CC      TRANSFERASE: Tyrosine-protein kinase; ATP-binding; Oncogene;
CC      KW      Glycoprotein; Phosphorylation.
CC      FT      DOMAIN 132 399
CC      FT      NP BIND 138 146 ATP (BY SIMILARITY).
CC      FT      BINDING 165 165 ATP (BY SIMILARITY).
CC      FT      ACT SITE 257 257 BY SIMILARITY.
CC      SQ      SEQUENCE 634 AA; 70891 MW; E705E33A0B01FCC CRC64;

```

Query Match: 24.5%; Score 1672.5; DB 1; Length 634;  
 Best Local Similarity: 50.8%; Pred. No. 3.9e-83;  
 Matches 359; Conservative 77; Mismatches 147; Indels 123; Gaps 17;

```

Qy      587 CAHYDPPFCVARGCGYKXPDLSTVYPIKPFDEBEGACQCPPTNCHSVVDLDDKCPAQEQ 646
Db      3 CAHFDGPRCVACAGVAGENDTL-VKVIADANAACVQLCHPNCRTGCKGGLBCP--- 58
Qy      647 RASPLTSIVSAV--GILLVVLGVVFGILLFNNFTV--SFWLRVPKVSASHLEPLTPSGAM 704
Db      59 NGSKTPSIAGVAGVGLCLVYVVGIGILYLRHRHIVRKRTLRLIQERELVEPLTPSGEA 118
Qy      705 PNOQAMRIKETELKAYVYVSSGAGTYRKGINIPDGNNAKIPVAIKVLRNTPSKANKE 764
Db      119 PNOAHLRIKETEFKVVVLGSGAGTVYKGLMPEGSKVPIPAIKELRATSPKAKE 178
Qy      765 ILDEAYVAVAGVSPYVSRLLGICLTSTVQVLTQVLTQMPYGCILLDHYRENRKGLSGODLLNMC 824
Db      179 ILDEAYVAVASVDNPRVCHLLGICLTSTVQVLTQVLTQMPYGCILLDHYREHNDNGSYLLNMC 238
Qy      825 MOIAKMSYLEDVLYHRDLAARVLYVKSPPNHVKITDFGLARLDIDETEVHADGKVP 884

```

RESULT 12  
 ERBB ALV  
 ID ERBB ALV STANDARD; PRT; 634 AA.  
 AC P00534;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).  
 GN V-ERBB.

```

Db      239 VOIAKMNYLEERLVRDLAARNVVKTPQHVKTIDFGLAKLGLADEKEVHAEGSKVPI 298
Qy      885 KMALESILRRFRTHOSDVWSYGVYWEMLMTFGAKPYDGIIPAREIDLLEKGERLPQPI 944
Db      299 KMALESILRRFRTHOSDVWSYGVYWEMLMTFGAKPYDGIIPAREIDLLEKGERLPQPI 358
Qy      945 CTIDVYMIWKCMWIDSECPRRRELVSSESMARDPORFVYIQ-NEDLGASPLDSTFY 1003
Db      359 CTIDVYMIWKCMWIDSECPRRRELVSSESMARDPORFVYIQ-NEDLGASPLDSTFY 418
Qy      1004 RSLLEDDMGDLVDAEYLVPOOGFCPPDAPAGAGVHHRSSSTRSGGDLTLGLEP 1063
Db      419 RSLLEDDMGDLVDAEYLVPOOGFCPPDAPAGAGVHHRSSSTRSGGDLTLGLEP 449
Qy      1064 SEEBAPRSP-----APSEGAGSDVDPDGLMGAKGLSLPHTDPSPLORYSEDPVLI 1118
Db      450 -----SRPLSLSLATSNNSATNCID-----RNGQGHPRFEDSFVQRYSSDPTGNF 496
Qy      1119 PSET--DGYVAPLTCSPQPEYVNPQDVRPQPSPREGLPAPAPAGATLERAKTSLPGKN 1176
Db      497 LEESIDGFL-----PAREVYNO--LMPKKPS-----TAMVNO 527
Qy      1177 GVYKDVF-----AFGAVENPEYLTPOGGAPOPHPPAFSPAFDNLVY 1221
Db      528 QIYNNISITAIKLPMDRYQNSHSTAVDNPEYL-----NTNQSPLAKTVFESSAPY 579
Qy      1222 DQ-----DPE-----RGAPSTFKGTPTAENPEYLGDPV 1252
Db      580 IQSGNHQINDPNPDYQODFLPNETKPNGLKVPAAENPEYLRVAAP 625

RESULT 13
ERBB_AVIER STANDARD: PRT: 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain Es4).
OC Viruses; Retroviruses; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN RP SEQUENCE FROM N.A. [1]
RC STRAIN=H;
RX MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
   gene family."
RL Cell 35:71-76(1983).
RN RP SEQUENCE OF 1-152 FROM N.A. [2]
RX MEDLINE=8423957; PubMed=6328658;
RA Dejure B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
   Saule S., Martin P., Stehelin D.;
RT "Sequencing the erbA gene of avian erythroblastosis virus reveals a
   new type of oncogene."
RL Science 224:1456-1459(1984).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
   tyrosine phosphate.
CC -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
   ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
   IN CHICKENS.
CC -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
   RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
   between the Swiss Institute of Bioinformatics and the EMBL outstation -
   the European Bioinformatics Institute. There are no restrictions on its
   use by non-profit institutions as long as its content is in no way

```

```

CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL, K02006; AAA42394.1; ALT_INIT.
DR EMBL, K01216; AAA42400.1; -.
DR PIR, A00644; TVYH.
DR HSSP, P11362; IEGK.
DR InterPro, IPR000719; Euk_Pkinase.
DR InterPro, IPR001245; Tyr_Pkinase.
DR Pfam, PF00069; Pkinase.1.
DR ProDom, PD000001; Euk_Pkinase.1.
DR SMART, SM00219; Tyrc.1.
DR PROSITE, PS00107; PROTEIN KINASE ATP.1.
DR PROSITE, PS00109; PROTEIN KINASE TYR.1.
DR PROSITE, PS00011; PROTEIN_KINASE_DOM.1.
KW Transferrase; Tyrosine-protein kinase; ATP-binding; Oncogene;
KW Glycoprotein; Phosphorylation.
FT DOMAIN 132 399
FT NP_BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT_SITE 257 257 BY SIMILARITY.
FT CONFLICT 29 29 R -> W (IN REF. 2).
FT CONFLICT 140 140 S -> F (IN REF. 2).
FT CONFLICT 146 146 I -> V (IN REF. 2).
SQ SEQUENCE 604 AA; 67633 MW; 76EBDD067450609 CRC64;

Query Match 23.9%; Score 1626; DB 1; Length 604;
Best Local Similarity 50.7%; Pred. No. 1,28-80;
Matches 349; Conservative 74; Mismatches 140; Indels 126; Gaps 16;

Qy 587 CAHYKDPFCVACRCSGVKPDLSYPMKFPDEBACQPCPINCHTSCYVDLDDKCPBAQ 646
Db 3 CAHFIDGPHCVACAGVAGENDTL-VKRYADANAVCOLCHNCRTCKGKGPGLGECF 58
Qy 647 RASPLSTVSAVY-GILLVWLVGVVFGILFENNFTV-SFWLRVPKVASHLPTPSGAM 704
Db 59 NSKTPSLAGVYGGILCLVVGILGLYLRNRHLYVKRTLRLLQGERLVPLTPSGA 118
Qy 705 PQOQMRILKETELRKVYKLGSGAFGYKGIWIDGENVKIPVAIKVLRNTSPYANKE 764
Db 119 PQOAMRLIKETEFKVKYLGSGAFGTIKGIMIEGEVYKIPVAIKELRATSPYANKE 178
Qy 765 ILDEAVYAGVSPYSLIGICTSTVQLVQLMPYGGILDHVENRGLSODPLNMC 824
Db 179 ILDEAVYASVNPVHCRLLGICLSTVQLVQLMPYGGILDHVENRGLSODPLNMC 238
Qy 825 MOIAKMSYLEDRVLRDLAARNVVKSPNHVKITDFGLARLLDIDETRYADGKVP 884
Db 239 VOIAKMNYLEERLVRDLAARNVVKTPQHVKTIDFGLAKLGLADEKEVHAEGSKVPI 298
Qy 885 KMALESILRRFRTHOSDVWSYGVYWEMLMTFGAKPYDGIIPAREIDLLEKGERLPQPI 944
Db 299 KMALESILRRFRTHOSDVWSYGVYWEMLMTFGAKPYDGIIPAREIDLLEKGERLPQPI 358
Qy 945 CTIDVYMIWKCMWIDSECPRRRELVSSESMARDPORFVYIQ-NEDLGASPLDSTFY 1003
Db 359 CTIDVYMIWKCMWIDSECPRRRELVSSESMARDPORFVYIQ-NEDLGASPLDSTFY 418
Qy 1004 RSLLEDDMGDLVDAEYLVPOOGFCPPDAPAGAGVHHRSSSTRSGGDLTLGLEP 1063
Db 419 RSLLEDDMGDLVDAEYLVPOOGFCPPDAPAGAGVHHRSSSTRSGGDLTLGLEP 449
Qy 1064 SEEBAPRSP-----APSEGAGSDVDPDGLMGAKGLSLPHTDPSPLORYSEDPVLI 1118
Db 450 -----SRPLSLSLATSNNSATNCID-----RNGQGHPRFEDSFVQRYSSDPTGNF 496
Qy 1119 PSET--DGYVAPLTCSPQPEYVNPQDVRPQPSPREGLPAPAPAGATLERAKTSLPGKN 1176
Db 497 LEESIDGFL-----PAREVYNO--LMPKKSTAM----- 524
Qy 1177 GVYKDVFAF-----GAVAVENPEYLTPOGGAPOPHPPAFSPAFDNLVY 1221

```

DB 525 -VQNGIYNFISLTLAISKLPMDSRYKQNSHSAVDNPEYL-----NTNQSLAKTVES 575  
 QY 1218 LYYWDDPPREGAPSTFKGTPTAENPEY 1246  
 DB 576 SPYWIQSGNHQ-----INLDNDPY 594

RESULT 14  
 ERBB AVIEU STANDARD; PRT; 540 AA.  
 ID ERBB\_AVIEU STANDARD; PRT; 540 AA.  
 AC P11273;  
 DT 01-JUL-1989 (Rel. 11, Created)  
 DT 01-JUL-1989 (Rel. 11, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 GN Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).  
 OS Avian erythroblastosis virus (strain ts167).  
 OC Viruses; Retroviridae; Avian type C retroviruses.  
 OX NCBI\_TaxID=103898;  
 RN [1]  
 RP MEDLINE=87064456; PubMed=2878364;  
 RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;  
 RT "A single amino acid substitution in v-erbB confers a thermolabile  
 RT phenotype to ts167 avian erythroblastosis virus--transformed erythroid  
 RT cells.";  
 RL Mol. Cell. Biol. 6:1751-1759(1986).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC -1- tyrosine phosphate.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC -----  
 CC EMBL: M31179; AAA42401.1; -  
 DR PIR: A25231; TVFEB.  
 DR HSSP: P11362; 1FGK.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00219; TYKFC; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KM Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;  
 KW Glycoprotein; Phosphorylation.  
 FT DOMAIN 132 399  
 FT NE\_BIND 138 146 ATP (BY SIMILARITY).  
 FT BINDING 165 165 ATP (BY SIMILARITY).  
 FT ACT\_SITE 257 257 BY SIMILARITY.  
 FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).  
 SQ SEQUENCE 540 AA; 60412 MW; 5B53297AA068B5D CRC64;

Query Match 22.8%; Score 1553; DB 1; Length 540;  
 Best Local Similarity 53.4%; Pred. No. 8; 8e-77;  
 Matches 330; Conservativity 67; Mismatches 131; Indels 90; Gaps 14;

QY 587 CAHYKDPFCVACPSGVKXDLSTYPMKPPDEEGACOPCPICVCHSCVDLDDKGPAPG 646  
 DB 3 CAHFDGPHCVKACPSAGVGEENTL-VWKYADANAVCOLCHPCTGCGGPGEGCP--- 58  
 QY 647 RASPLTISYSAVV-GILLVVIGVGFIIENFTV-SFWLRVKSASASLEPLTSGAM 704  
 DB 59 NSGKTPSIAGVGVGGLICLVVGLIGLYLRHRHIVKRTLRLLDERELVEPLTSGBA 118  
 QY PNAQWRILKETELRVKVLKSGAFGVYKGIPIPGENVKIPVAIKVLRNTSPANK 764

DB 119 PNAQWRILKETELRVKVLKSGAFGVYKGLMIPEGEKVTIPVAIKELREATSPANK 178  
 QY 765 ILDEAYVAGVSGPYVSRRLGICITSTYQVLTQMPFGCLLDHRENRGRIGSQDLLNC 824  
 DB 179 ILDEAYVAGVADNPVCCVLLGICITSTYQVLTQMLPGCLLDYIREHNDNGSOYLLWC 238  
 QY MOIAKMSYLEDAVLVERDILAAANVLKSPNNHVTITPGLARLLIDETEHADGKVP 884  
 DB 239 VOIAKMSYLEERHVRDILAAANVLKTPGHVITITPGLAKOLGADKEHAEGKVP 298  
 QY 885 KMALESILRRRFTQSDVMSYGVTELMFGAKPYDIPAREIPDLEKGERLPDP 944  
 DB 299 KMALESILHRITVQSDVMSYGVTELMFGAKPYDIPAREIPDLEKGERLPDP 358  
 QY 945 CTIDVYMTVMCMWIDSCRPREFEIVSEFRMRDPPRFVVIC-NEDLGPASPLDSFY 1003  
 DB 359 CTIDVYMTVMCMWIDSCRPREFEIVSEFRMRDPPRFVVIC-NEDLGPASPLDSFY 418  
 QY 1004 RSLLEDDMDGLVDAEYLVPQGFPCDPAPAGAGVHHRSSSTRSGGDLTLGLEP 1063  
 DB 419 RTIMEEDMEDIVDADEVLPHQGF-----NSPST----- 449  
 QY 1064 SEEPAPRSP-----APEEGAGSDVFQDGLMGAKGLQSLPTHDPSPLOKYSDDPTVPL 1118  
 DB 450 -----SRTPLLSSATSNNSATNCIDNNG-----H----- 476  
 QY 1119 PSETDGYVAPLTPSGPQPEYVQNPDPQPSPREGPPAPAPAGAT-LEAKTLSPGNG 1177  
 DB 477 PVREDGFL-----PAPRYVNG--LMPKRPSTANVQNVYISITLSKLPIDSRYN- 527

QY 1178 VKQVFAFGAVENPEYL 1195  
 DB 528 -----SHSTAVDNEPYL 539

RESULT 15  
 EGFR\_CHICK  
 ID EGFR\_CHICK STANDARD; PRT; 703 AA.  
 AC P13387;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 GN Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER  
 GN (Fragment).  
 GN EGFR.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP MEDLINE=88261272; PubMed=3260329;  
 RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,  
 RA Ullrich A., Vamvakis B., Schlessinger J., Givol D.;  
 RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,  
 RT expression in mouse cells, and differential binding of EGF and  
 RT transforming growth factor alpha.";  
 RL Mol. Cell. Biol. 8:1970-1978(1988).  
 CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,  
 CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND  
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC -1- tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; M20386; AAA48760.1; -  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR01245; Tyr\_pkinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR SMART; SM00261; FU\_4.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; PARTIAL.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; PARTIAL.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; PARTIAL.  
 DR Transmembrane; Glycoprotein; Receptor; Signal; Transferase;  
 KM Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
 FT SIGNAL 1 30  
 FT CHAIN 31  
 FT DOMAIN 31 654  
 FT TRANSSEM 655 667  
 FT DOMAIN 668 >703  
 FT DISULFID 197 206  
 FT DISULFID 201 214  
 FT DISULFID 222 230  
 FT DISULFID 226 238  
 FT DISULFID 239 247  
 FT DISULFID 243 255  
 FT DISULFID 258 267  
 FT DISULFID 271 298  
 FT DISULFID 302 314  
 FT DISULFID 318 333  
 FT DISULFID 336 340  
 FT DISULFID 513 522  
 FT DISULFID 517 530  
 FT DISULFID 533 542  
 FT DISULFID 546 562  
 FT DISULFID 565 581  
 FT DISULFID 569 589  
 FT DISULFID 592 601  
 FT DISULFID 605 627  
 FT DISULFID 630 638  
 FT DISULFID 634 646  
 FT CARBOHYD 134 134  
 FT CARBOHYD 190 190  
 FT CARBOHYD 200 200  
 FT CARBOHYD 359 359  
 FT CARBOHYD 368 368  
 FT CARBOHYD 420 420  
 FT CARBOHYD 573 573  
 FT CARBOHYD 578 578  
 FT CARBOHYD 613 613  
 FT CARBOHYD 633 633  
 FT CARBOHYD 648 648  
 FT NON\_TER 703  
 SQ SEQUENCE 703 AA; 77427 MW; AEF2DE11B735A690 CRC64;

Query Match 22.3%; Score 1518; DB 1; Length 703;  
 Best Local Similarity 43.1%; Pred. No. 9.1e-75;  
 Matches 305; Conservative 109; Mismatches 263; Indels 30; Gaps 14;

QY 8 RWGLLALPPGAA-----STVCTGDMKRLRPASPEHLDMLRLHYGGCCVYQGNLE 61  
 DB 13 RGAALVLLLVGLGSAVEEKVCGQNNKLTQLGHVEDHFTSLQRRYNNCEVALSNLE 72  
 QY 62 LTYLPTNASLSFLQIDQVQGVLIHNNQVQVPLQRLIVRGTOLEFEDNYALAVLDNGD 121  
 DB 73 ITVEHNDLTFLEKTIQEVAGVLIHNNQVQVPLQRLIVRGTOLEFEDNYALAVLDNGD 132  
 QY 122 PLNNTTPTGASPGRLRLQRLSLTEILKGVLIQRNPQLCYQDTIILKDIFFHKNQDLAL 181

DB 133 -MNTQ-----GLRELPMKRLSELINGVXISNNPKLCNMOTVLANDIIDTSRK-PL 182  
 QY 162 TLID-TNRSRACHPCSPMKCKSRCKSESSDCCQSLTRVCGGGA-RKSGELPTDCQHO 239  
 DB 183 TVLDFASVSSCPKCHPCTEDHCHWGAECQNCOTLYKVICAOQCSGRGRGVPSDCCQHO 242  
 QY 240 CAAGCTGPKHSDCLACLFHNSGICELCPALVTYNTDTFFSMPREGRYTFGASCVTAC 299  
 DB 243 CAAGCTGPRESDCLACRFRRDATCKCTCPFLVLYNPTTYGMDVNVBEKSTGATCVNEC 302  
 QY 300 PNYLSTDVSGCTVCPHANOVTADGTQRCCKSPCARCYGLGMOYIKANSKFTGI 359  
 DB 303 PHNYVYTDHSGCVRSCNNTDYEY--EENGVRCCKKDDGLCSKVCNIGIGELGLIIS-INA 360  
 QY 360 TELF-PACCKTFGSLAPSPESFGDPSANTAPLOPELOVFEETLEETGYLYISAMPDS 418  
 DB 361 TNIDSFCKCTKINGDVSILPVAFLGDARTKLPDPKCLDVFRYVKEISGFLLIQAMPDN 420  
 QY 419 LPDLVFOQLQYIRGRILHNGAVSLTLOGLSIWLGLSLRLSRLSGLALIHNTHLCFVH 478  
 DB 421 ATDLVAFERLEIRGRTRQHQGYSLAVNVLKIQSLKLSKEISDGLAIIMKNKMLCYAD 480  
 QY 479 TYPMQQLFRNPQALLHTANRPEDCEVGEGLACHQLCARHGMGSPQCVNCSQFLQG 538  
 DB 481 TNMWSLFPATQSQKTKIIONRNKNDCTADRHVCDPLCSDVCGMGSPHCFSCRFPSROK 540  
 QY 539 ECVEEGRVYQGLFREYVNAHCLPCHPECPQNG---SVTCFPEADQCVACAHYKDPFF 595  
 DB 541 ECVKQCNILIQSPREFRDSKLPCHSEGLVNSTAYNTTSGPGRDHMKCAHFIDGPH 600  
 QY 596 CVARCPGVPKPLSTMPKPKFPDEGACQPCPINCCTHSCVDLDXGCPAEQASPLTSIV 655  
 DB 601 CVACPGAGVLGNDPL-VKYNADANAVQCLPNCCTRGCKGGLGCP---NGSKTPISA 656  
 QY 656 SAVV-GILVVVLGVVFGILIFNFTV-SFMLRVPKVSASHLEPLTP 700  
 DB 657 AGVGGLLCLVAVGIGILYLRRIIVRKTRILQERELVEPLTP 703

Search completed: July 22, 2003, 09:18:21  
 Job time : 20.625 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.75 Seconds

(without alignments)  
5522.503 Million cell updates/sec

Title: SEQ-653-675-14

Perfect score: 6814  
Sequence: 1 METALCRMGILLALIPGA.....TFKGTPTAENPEYGLDVPV 1253

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL\_21:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp Vertebrate:\*
- 14: sp Unclassified:\*
- 15: sp\_virus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6097	89.5	1259	6 O18735	O18735 canis fam1
2	3060	44.9	1209	11 Q9GX70	Q9GX70 ractus nov
3	3038	44.6	1210	11 Q9EP98	Q9EP98 mus musculu
4	2647	38.8	1165	13 Q9YH40	Q9YH40 xiphophorus
5	2636.5	38.7	1137	13 Q9W6F6	Q9W6F6 gallus gall
6	2265	33.2	1338	13 P79754	P79754 fugu rubrip
7	2012.5	29.5	1433	5 Q9B1H9	Q9B1H9 anopheles g
8	1871	27.5	419	4 Q9UK79	Q9UK79 homo sapien
9	1739	25.1	357	11 Q8R2X1	Q8R2X1 mus musculu
10	1697.5	24.9	412	4 Q8WYV0	Q8WYV0 homo sapien
11	1643	24.1	729	15 Q86712	Q86712 avian rous-
12	1641	24.1	567	15 Q86714	Q86714 avian rous-
13	1576.5	23.0	962	15 Q64895	Q64895 avian eryth
14	1568	23.0	545	15 Q85468	Q85468 avian eryth
15	1506.5	22.1	655	11 Q9WVF5	Q9WVF5 mus musculu
16	1490.5	21.9	643	11 Q9ERV6	Q9ERV6 mus musculu

17	1268	18.6	1193	5 Q9Y1X8	Q9Y1X8 ephydactia f
18	1180.5	17.3	1368	5 Q23821	Q23821 caenorhabdi
19	1142	16.8	1717	5 Q26566	Q26566 schistosoma
20	1126	16.5	527	13 Q90836	Q90836 gallus gall
21	1001.5	14.7	478	11 Q9ESE0	Q9ESE0 ractus nov
22	942.5	13.8	599	13 Q9PSH2	Q9PSH2 gallus gall
23	906	13.3	165	4 Q14256	Q14256 homo sapien
24	887	13.0	176	11 Q923V5	Q923V5 ractus nov
25	806.5	11.8	346	13 P11776	P11776 xiphophorus
26	778	11.4	435	5 Q8SZW1	Q8SZW1 dirosophila
27	754.5	11.1	311	13 Q99162	Q99162 xiphophorus
28	741.5	10.9	1362	13 Q9PVZ4	Q9PVZ4 xenopus lae
29	734	10.8	331	4 Q9BDU7	Q9BDU7 homo sapien
30	725	10.6	1671	5 Q9NJV5	Q9NJV5 biotomphalari
31	723	10.6	149	6 Q9BG66	Q9BG66 oryctolagus
32	680.5	10.0	1368	13 Q8UW85	Q8UW85 paratichthy
33	680	10.0	1418	13 Q93457	Q93457 scophthalmu
34	673.5	9.9	1369	13 Q8UW86	Q8UW86 paratichthy
35	672	9.9	1412	13 Q8UW84	Q8UW84 paratichthy
36	660.5	9.7	1472	5 Q9U5A8	Q9U5A8 bombyx mori
37	658	9.7	1358	13 Q73798	Q73798 xenopus lae
38	640	9.4	1245	13 Q9YGR8	Q9YGR8 scophthalmu
39	636.5	9.3	1418	13 Q8UW83	Q8UW83 paratichthy
40	630.5	9.3	2144	5 Q9VD94	Q9VD94 dirosophila
41	624	9.2	1371	11 Q9QVW4	Q9QVW4 ractus sp.
42	602.5	8.8	1091	4 Q9UMQ4	Q9UMQ4 homo sapien
43	595.5	8.7	1072	4 Q9BT80	Q9BT80 mus musculu
44	595	8.7	987	11 Q91YV0	Q91YV0 mus musculu
45	586	8.6	1036	4 Q07912	Q07912 homo sapien

## ALIGNMENTS

### RESULT 1

O18735 PRELIMINARY: PRT: 1259 AA.

AC O18735; 01-JAN-1998 (TREMBLrel. 05, Created)  
 DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)  
 DR 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Erbs-2.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Yokota H.;  
 RT Submitted (Oct-1997) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB008451; BAA2127.1; -.  
 DR HSP; P11362; 1FGK.  
 DR Interpro: IPR002048; EF-hand.  
 DR Interpro: IPR000494; EGFR\_L domain.  
 DR Interpro: IPR000719; Euk\_Pkinase.  
 DR Interpro: IPR002174; Furin-like.  
 DR Interpro: IPR001245; Tyr\_Pkinase.  
 DR Interpro: IPR004019; YIP\_motif.  
 DR Pfam: PF00157; Furin-like; 1.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR Pfam: PF02757; YIP\_2.  
 DR Prodom: PDOM00001; Euk\_Pkinase; 1.  
 DR SMART; SMO0261; FU; 3.  
 DR SMART; SMO0219; Tyrc; 1.  
 DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Transferase; Tyrosine-protein kinase  
 SQ SEQUENCE 1259 AA; 137989 MW; E3764D94C4CD46 CRC64;

Query Match 89.5%; Score 6097; DB 6; Length 1259;  
 Best Local Similarity 89.1%; Pred. No. 0;  
 Matches 1126; Conservative 48; Mismatches 74; Indels 16; Gaps 4;

1 MELAALCRWGLLLLLPPGAASVCTGTGTWKLRLPSPETHLDMHLXOGGVVQGNL 60  
 1 MELAAWCGWGLLLLLPSGAAGTGVCTGTWKLRLPSPETHLDMHLXOGGVVQGNL 60  
 61 ELTYLPTNASISFLQDIOEVGYVLIANOVYQVLIQRLIRVGTQLFEDNYALAVLDNG 120  
 61 ELTYLPANASISFLQDIOEVGYVLIANOVYQVLIQRLIRVGTQLFEDNYALAVLDNG 120  
 121 DLNNTPTVTGASGGARELOSLSTELIKGVLIQGNPOLCYODTILMDIFHXNOLA 180  
 121 DLEGGIIPAPGAAGGARELOSLSTELIKGVLIQGNPOLCYODTILMDIFHXNOLA 180  
 181 LTLIDTNRBACHPCSPWCKGRCWGESSEDDQSLRTVTCAGGACRCKGRLPTDCHEQC 240  
 181 LTLIDTNRBACHPCSPWCKGRCWGESSEDDQSLRTVTCAGGACRCKGRLPTDCHEQC 240  
 241 AAGCTGPHNSDCLACIHNHSGICELHCPALVYNTDTFESMPREGRTTGGASCTYACP 300  
 241 AAGCTGPHNSDCLACIHNHSGICELHCPALVYNTDTFESMPREGRTTGGASCTYACP 300  
 301 YNYLSTDVGSCTLVCPLHNOEVTAEADGTORCEKSKPCARCYGLGMQYIKANKRFGIT 360  
 301 YNYLSTDVGSCTLVCPLHNOEVTAEADGTORCEKSKPCARCYGLGMQYIKANKRFGIT 360  
 361 ELERAGCKKITGSLAFLEPESFDGPASTAPLOEQLOVFTLEITGYLYISAMPDLP 420  
 361 IQEFAGCKKITGSLAFLEPESFDGPASTAPLOEQLOVFTLEITGYLYISAMPDLP 420  
 421 DLSEFONLOVIRGILHNGAYSILFLOGISMLGRLSRELGSALILHNTILCFPHTV 480  
 421 NLSEFONLOVIRGILHNGAYSILFLOGISMLGRLSRELGSALILHNTILCFPHTV 480  
 481 PWDOLFERNPHOALLHSANRPEECVGEGLACYP-CANHCHGMPPTCQVNCSCPLRQEC 539  
 481 PWDOLFERNPHOALLHSANRPEECVGEGLACYP-CANHCHGMPPTCQVNCSCPLRQEC 539  
 541 VEEGRVLOGLEPREVYNAHCLPCHPECQPNQSVTCFPEPDQVACAHYNDPFCYARC 600  
 541 VEEGRVLOGLEPREVYNAHCLPCHPECQPNQSVTCFPEPDQVACAHYNDPFCYARC 600  
 601 PSQVPCDLSFPIWKFADDEGTCQPCPINCTHSCADLDEKGCAPRASTPVTIIAAVVG 659  
 601 PSQVPCDLSFPIWKFADDEGTCQPCPINCTHSCADLDEKGCAPRASTPVTIIAAVVG 659  
 661 ILLVVLGVVFGILLI-----FNNFTVSFMLRVKVSASHLEPLTSGAMNQNOMRLK 714  
 661 ILLVVLGVVFGILLI-----FNNFTVSFMLRVKVSASHLEPLTSGAMNQNOMRLK 714  
 715 ETELRKVKVLGSGAFIYKGIWIPDGENVKIPVAIKYLRENTSPKANKETLDAVYVAG 774  
 715 ETELRKVKVLGSGAFIYKGIWIPDGENVKIPVAIKYLRENTSPKANKETLDAVYVAG 774  
 716 ETELRKVKVLGSGAFIYKGIWIPDGENVKIPVAIKYLRENTSPKANKETLDAVYVAG 775  
 716 ETELRKVKVLGSGAFIYKGIWIPDGENVKIPVAIKYLRENTSPKANKETLDAVYVAG 775  
 775 VGSFVSVLLGICLTSTVOLTQOLMPYGCILDHRENKRGISQDILNMCQIAKMSYL 834  
 775 VGSFVSVLLGICLTSTVOLTQOLMPYGCILDHRENKRGISQDILNMCQIAKMSYL 834  
 776 VGSFVSVLLGICLTSTVOLTQOLMPYGCILDHRENKRGISQDILNMCQIAKMSYL 835  
 776 VGSFVSVLLGICLTSTVOLTQOLMPYGCILDHRENKRGISQDILNMCQIAKMSYL 835  
 835 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEYHAAGKVPKIMWLESILR 894  
 835 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEYHAAGKVPKIMWLESILR 894  
 836 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEYHAAGKVPKIMWLESILR 895  
 836 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEYHAAGKVPKIMWLESILR 895  
 885 RRFTHQSDVMSGYVTVMELTFGAKPYGIPAREIPDLLEKGERLPORPITTDIVYIMV 954  
 885 RRFTHQSDVMSGYVTVMELTFGAKPYGIPAREIPDLLEKGERLPORPITTDIVYIMV 954  
 886 RRFTHQSDVMSGYVTVMELTFGAKPYGIPAREIPDLLEKGERLPORPITTDIVYIMV 955  
 886 RRFTHQSDVMSGYVTVMELTFGAKPYGIPAREIPDLLEKGERLPORPITTDIVYIMV 955  
 955 KCMIDSECRPRFRELVESEFRMAKDPORFVYIQNEDGASPSPLDSTYRSLLDDDDMGD 1014  
 955 KCMIDSECRPRFRELVESEFRMAKDPORFVYIQNEDGASPSPLDSTYRSLLDDDDMGD 1014  
 956 KCMIDSECRPRFRELVESEFRMAKDPORFVYIQNEDGASPSPLDSTYRSLLDDDDMGD 1015  
 956 KCMIDSECRPRFRELVESEFRMAKDPORFVYIQNEDGASPSPLDSTYRSLLDDDDMGD 1015  
 1015 LVDABEYLVPOQGFPCPDPAAGAGMTHRRSSSTRSGGSDTLTLGLEPSEEEARPSPLA 1074  
 1015 LVDABEYLVPOQGFPCPDPAAGAGMTHRRSSSTRSGGSDTLTLGLEPSEEEARPSPLA 1074

1016 LVDABEYLVPOQGFPCPDPAAGAGMTHRRSSSTRSGGSDTLTLGLEPSEEEARPSPLA 1075  
 1075 PSEGASDVFDGDLGGAAGKQSLPTHDPSFLQRYSDPTVPLPSETDGYVAPLTGSPQ 1134  
 1076 PSEGASDVFDGDLGGAAGKQSLPTHDPSFLQRYSDPTVPLPSETDGYVAPLTGSPQ 1135  
 1135 PEYVNOPEVWPQPPALBEPPLPSPRAGATLERPKTILSPKILSPGNVGVKDFAFGSAV 1195  
 1136 PEYVNOPEVWPQPPALBEPPLPSPRAGATLERPKTILSPKILSPGNVGVKDFAFGSAV 1195  
 1190 ENPEYLTPOGGAAPOPHPPPAFSPAFDNLYYWDOPPERGAPSTFGPTAENPEYLT 1249  
 1196 ENPEYLTPOGGAAPOPHPPPAFSPAFDNLYYWDOPPERGAPSTFGPTAENPEYLT 1255  
 1250 DVPV 1253  
 1256 DVPV 1259

RESULT 2  
 Q9QX70 PRELIMINARY; PRT; 1209 AA.  
 ID Q9QX70  
 AC Q9QX70;  
 DT 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor.  
 GN BGRF.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RX MEDLINE=90258888; PubMed=2342456;  
 RA Petch L.A.; Harris J.; Raymond V.W.; Blaehand A.J.; Lee D.C.;  
 RA Earp H.S.;  
 RT "A truncated, secreted form of the epidermal growth factor receptor is encoded by an alternatively spliced transcript in normal rat tissue.";  
 RL Mol. Cell. Biol. 10:2973-2982 (1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RA Petch L.A.;  
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RA Guttridge K.; Dawson T.L.; Earp H.S.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; M37394; AAF14008.1; -.  
 DR HSSP; P11362; 1FGK.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR InterPro; IPR002179; Euk\_Pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_Pkinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF01030; Recept\_L domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PDOM0001; Euk\_Pkinase; 1.  
 DR SMART; SM00261; Fur; 3.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE\_ATP; 1.  
 DR PROSITE; PS00111; PROTEIN KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE\_TYR; 1.  
 KM ATP-binding; Receptor; Transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 1209 AA; 134891 MW; 96FEE7FECB1B773 CRC64;

Query Match 44.9%; Score 3060; DB 11; Length 1209;  
 Best Local Similarity 49.3%; Pred. No. 8; 9e-223;

Matches 630; Conservative 164; Mismatches 366; Indels 118; Gaps 28;

QY 3 LALACRWGLLALPFGA-ASTOYCTGTDMLRLPAPERTHMLRLHYGGCVGNNE 61  
 Db 15 LALCAAG-----GALBEKVCQSTNRILTQGFEDHFLSLQMFNCCVVLGNLE 66  
 QY 62 LTYLPTNASLSFLQDIOEVQGVYLIANQVRQVPLQRLIRVSTQLEFQVYALAVLNDG 121  
 Db 67 ITVQRNYDLSFLKTIQEVAGVYLIANTVERIPLENIQIRGNALYENTYALAVLSN-- 124  
 QY 122 PLNNTPTVTGASPEGJRELOLRSLTELKGVLIQRPOLCYDDTLMKRIFKHNQML 181  
 Db 125 -----YGTNKTGLRELPRLNQLQELIGAVRSPNPILCNMTIQMRDLY-QDVFLS 175  
 QY 182 TLIDTNRS-RACHPCSPMKSGSRGWSSESDCCSLTRTVGAGCA-RCKGRLPTDCEH 239  
 Db 176 MSMDVQRHLTGCPKCDPSCPNGSCWGRGECNQLTITICAGCCSRRCGRSPSDCHQ 235  
 QY 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALTYNTDTESMPNPEGRTFQASCYTAC 299  
 Db 236 CAAGCTGPRSDCLVCHRFDEATCTKTCPLMLYNTTYQMDVNPBGKXSFQATCVKCC 295  
 QY 300 PYNLSTDVGSCTLVCLHNOEVTAEQGTQRCCKSPCARVCYGLGMQYIKANSKFIGI 359  
 Db 296 PRNVVTDHSGCVRACGPDYEV-EDGVSCKCKKCGPCRCXVNGIGISFK-DTSLINA 353  
 QY 360 TELE-FACCKKITGSALFLPESPDGPASTAPLOEQLOVETLEITGYLYISMPDS 418  
 Db 354 TIKHFKYCTAISGLDILFLVAFKGSFTRTPRLDELEILITVKEITGLFIQAMPEN 413  
 QY 419 LPDLVSFQNTQVTRGRILHNGAVSLTQGLISMLGRSLREKSGSLALHHTHCIFYH 478  
 Db 414 WTDLHAFENLEITRGRKQHGQSLAVAGLNTSLGRSKETISDGVIIISGRNLCYAN 473  
 QY 479 TVPMQDLFRNPHQALLHTANRPDECVBEGSLACHQICARHCWGPPTQCVNCSQRLRQ 538  
 Db 474 TIKWKLFGTPNCKTINNRARCKATNHCNPICSSBSCGPEPTDCVSGQVNSRGR 533  
 QY 539 ECVGEGRVLOGLPREYVNAHCLPCHPECPONGSVTCPEPADQCAACHYDPRPCVA 598  
 Db 534 ECVDKCNLTGEBREPEFENSECQCHPECLPOTMNITCTGRGDNCKCKAHYVDGHCYK 593  
 QY 599 RCPGKVPDLSTYPIWKFPDEBAGACPPICTHSCVLDKCKPAEQRASPLTISIVA 657  
 Db 594 TCSSGIMGENNTL-VKMFADANNVCHLCHANCTYGGAGPGLKCC-QQPEGRPIELIANG 650  
 QY 658 VNGILLVNVGVVFGILIF--NNFTVSFPLRPKYSASHLEPLTSGAMPQACQRIILK 714  
 Db 651 IVGGLFLFIV-VALGIGLFMRRLQVARKTLRLLOERELVEPLTPSGEAPNQAHRLIK 709  
 QY 715 ETELARKVVLGSAFGTVYKGIWIPDENYKIFVAIKVLRENTSPANKETIDEAYVMAG 774  
 Db 710 ETEFKIKVLGSAFGTVYKGLMIPGEKIKIPVAILKELREALSPANKETIDEAYVMS 769  
 QY 775 VSGPYVSRLLIGTCLTSLVQLVLTQMPYGLLDHVRENRGLSGODLLNMCQIAKMSYL 834  
 Db 770 VDMPHVCRLIGTCLTSLVQLITQMPYGLLDVREHKNIGSQVLLNMCVQIAKMSYL 829  
 QY 835 EDVRLVHRDLAARNVLKSPNHVKITDPGLARLLDDEPYHADGKVPKXMALESLIR 894  
 Db 830 EDRRLVHRDLAARNVLKTPQHKITDPGLAKLLGAEKRYHAEGKVPKXMALESLIR 889  
 QY 895 RRTTHOSDWSYGVYTWELMTFGAKPYDGIIPASEIFDLLEKGERLPQPPICITDVMYV 954  
 Db 890 RIYTHOSDWSYGVYTWELMTFGSKPYDGIIPASEISILEKGERLPQPPICITDVMYV 949  
 QY 955 KCMWIDSECPREFRELVSFESRMAADPQRFVVIQ-NEDGPNASPLDSTYRSLEDDMG 1013  
 Db 950 KCMWIDSECPREFRELVSFESRMAADPQRFVVIQ-NEDGPNASPLDSTYRSLEDDMG 1009  
 QY 1014 DLVDAEEYLVPOGFFCPDPAPAGAGVHHRHSSSTRSGGDLTLGLPSEHEAPRSL 1073  
 Db 1010 DVVDADEYLVPOGFF-----NSPST-----SRTPL 1035

QY 1074 APEAGASPVFDGDLGMAKAGLQSLPTHPDEPLQRYSEDPVTPVPSER--OGVYAPLTC 1131  
 Db 1036 LSSLSANST-----SSTVACINNGSCRYKEDFLQRYSDPFSVLTEDNIDDTFL----- 1086  
 QY 1132 SPOPEYVQPDVRFPPSPREGPLPAPAPAGATLERAKTLSPCKNGVNVKVAFGAVEN 1191  
 Db 1087 -PVPEYVING-SVPRKSPAGSVQNPVYHNPFLH-----ABGRDLHYQN--PHSNVAVSN 1134  
 QY 1192 PEYVLT-TPQGAAPQCHPPAPSPAFDNLVYMPQ-----DP-----PERGAPST 1234  
 Db 1135 PEYVLTQAQ-----PTCLSSGFDSSALMTQKSHQMSLDNPYQDFFPKKAPNGI 1185  
 QY 1235 FKGTPTAENPEYVGLDVP 1252  
 Db 1186 FKGTPTAENPEYVGLDVP 1202

RESULT 3  
 Q9EP98  
 ID Q9EP98 PRELIMINARY; PRT; 1210 AA.  
 AC Q9EP98;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor isoform 1.  
 GN EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RC SEQUENCE FROM N.A.  
 RP STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 RA Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balseubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
 RA Mainle N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative Egr transcripts encoding truncated receptor  
 RT isoforms."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 RA Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balseubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,  
 RA Mainle N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative Egr transcripts encoding truncated receptor  
 RT isoforms."  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF275366; AAG28045.1; -  
 DR EMBL; AF275364; AAG28045.1; JOINED.  
 DR EMBL; AF275365; AAG28045.1; JOINED.  
 DR EMBL; AF275367; AAG24386.1; -  
 DR HSSP; p11362; 1FGK.  
 DR MGD; MGI:95294; Egr.  
 DR InterPro; IPR000345; CytC\_heme\_bind.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR01245; Tyr\_kinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Kinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00261; FU; 5.  
 DR SMART; SM00220; S\_TKC; 1.  
 DR SMART; SM00219; TYKIC; 1.

DR PROSITE, PS00190; CYTOCHROME C; UNKNOWN\_1.  
 DR PROSITE, PS00107; PROTEIN KINASE, ATP, 1.  
 DR PROSITE, PS00101; PROTEIN KINASE, DOM, 1.  
 DR PROSITE, PS00109; PROTEIN KINASE, TYR, 1.  
 DR ATP-binding; Receptor; Transferase.  
 KM SEQUENCE 1210 AA; 134840 MW; 62CD0021C9DE32E18 CRC64;

Query Match 44.6%; Score 3038; DB 11; Length 1210;  
 Best Local Similarity 48.9%; Pred. No. 4,1e-221;  
 Matches 622; Conservative 167; Mismatches 370; Indels 114; Gaps 25;

11 LLLALLPFGAA--STOVCTGDMKRLPASPETHLDMRLHYOCQCVQVQNGLETTYPTN 68  
 14 LTLTCAAGALAEKXKVCQGTNSNLTQGTEDHFLSLQRYNNCEVVLGNLETTYQRN 73  
 69 ASLSFLQDIEVQGVILAHNQVROVPLQRLIRYRGQLFEDNYALAVLDNGDPLNNTTP 128  
 74 YDLSFLTKIOEVAGVYLIALNTVERIPLENLQIIRGNALYENTYALAILSN----- 124  
 129 VTGASPGGLRELOLSLTELKGGVLIQORNPOLCYOPTILMKDI----FHKNQALATLI 184  
 125 -YGINRTLRRLPKNLDELIGAVRPSNNPILGNMTIQRDILVQVNFPSNMSMDL--- 180  
 185 DTNRSRACHPCSPMKCSRCWGESSEDCQSLTRTVCAAGCA-RCKGPLPTDCHEQCAAG 243  
 181 -QSHPSCPKCDPSCPNCSWGCGEHCQKLTIIICAQCSHRCGRSPSCCHNQCAAG 239  
 244 CTGPKHSDCLALPHNHSGLICELHCPALVTYNTDPEFMPBEGRYTFGASCVACPYNY 303  
 240 CTGPRESCLVQCKFODEATKCTCPPLMLYNTPTTQVDNVEKTSIGATCVCKCRNY 299  
 304 LSTDVSGCTLVCPHNDVETAEQGTORCEKSKPCARCYGLAQVQYKANSKFIGITELE 363  
 300 VVTHGSCVRAAGCPDYEV-EDGIRKCKCDGCRKCNIGIGEFK-DTLINATNINIK 357  
 364 -FAGCKTFGSLAPLPSFDDPASNTPALPEQLOVETLEETTYLYIYSANWDSLPLD 422  
 358 HFKYCTAISGDLHILPVAFKDSFTRPPLDPRELEIKTYKETLGLLQAWDDNNTDL 417  
 423 SVFQMLQVIRLHNGAVSLTLOGLISWLGSLRLSRLSGSLALHINTLHCVHTVPW 482  
 418 HAFENLEIRGRTRKHQGFSLAVGLNITSLGRSLKISGVDYIISGNRLCAVANTINW 477  
 483 DQFRNPQALHTANPEDECVGEGLAGHOLCARHGCGWGPPTQCNCVSGPLRGQCEVE 542  
 478 KKLFGTPQKTKIKNNRAEKXCKAVNHCNPLCSBEGGPEPDCVSCQVNSGRECEVE 537  
 543 ECRVLQGLPREVYANRHLCPHPECOFONGSVTCGPEADQCVACAHYKCDPPFCVACPS 602  
 538 KGNILEGPREFEVENSEICIQHPBELPQANMITCTGRSPDNCICCAHIDDPHCVKTCFA 597  
 603 GVKPDLSTMPWKPFDEBGACOPCPINCTHSCVDLDKGCFAEGRASPLTISVAVVGL 662  
 598 GIMGENNTL-VKXYADANNVCHLCHANCTYGCAGPGLQGCWVSPGKPIPIATGIQVGL 656  
 663 LVVYGVVFEGLIF---NNFTVSPFLRVKVSASHLEPLTSGAMPNQAORILKETEELR 719  
 657 LPIIV-VALGIGLFRKRRHIVRKRLRLLEBELVEPLTSGEAPNDARILKETEELR 715  
 720 KVKVLGSGAGFTVYKGINPDGENYKIPVAIKVIRENTSPRANKELIDEAVYMAVGSFY 779  
 716 KIKVLGSGAGFTVYKGLWIPGEKVKIPVAIKELREATSPRANKELIDEAVYMAVSVNPH 775  
 780 VSRLLGICLTSTVOVLQMLPQGLLDHVRNBRGLSGODLNNCMOIAKMSYLEVRL 839  
 776 VCRLLGICLTSTVOQLTQMLPQGLLDHVRNBRGLSGODLNNCMOIAKMSYLEVRL 835  
 840 VHRDLAANVAVKSNHVKITDFGLARLLDIDETRYADGKVPKMMALSTILRRFTH 899  
 836 VHRDLAANVAVKTPQHKITDFGLAKLGAEEKVEYHAGKVPKMMALSTILHRIYTH 895  
 900 OSDVWSYGVTVWELMTFGAKPYDGI-PAREIDPLEKGRLLPOPICTIDYVMVWKMMI 959

DB 896 OSDVWSYGVTVWELMTFGSKPYDGI-PASDISLLEKGRLEPOPICTIDYVMVWKMMI 955  
 QY 960 DECEPRFRELVSFSRMARDQRFVVIQ-NEDLCAPASPLSTFYRSLLEDMDGLDVA 1018  
 DB 956 DADSRKPFRELILLESKMARDDQRLVIVQGDGRMHLPSPTDNNFRLAMDEDMEDVYDA 1015  
 QY 1019 EEYLVQGGFPCFCDPAPAGAGVHHRRHSSSTRSGGDLTLLEPSEEPAPSPPLAPSEG 1078  
 DB 1016 DEYLVQGGF-----NSPST-----STPILLSLS 1041  
 QY 1079 AGSDVFDGLGMAAGKGLSLPTDPSPLQRYSEDPYVLPSET-DGYVAPLTSPOPE 1136  
 DB 1042 ATSN-----NSTVACINRNSCRVKEDAFQRYSSDPGTAVTEDNIDDAFL-----PVBE 1091  
 QY 1137 YVQPDYRQPPSPRPGPLPAPRPAATLERAKTISPGKNGVYKDVAFGAVENPEYL- 1195  
 DB 1092 YVQ--SVPRPAGSVQNPYVHNQPLHP-----APGDLHYQ--PNSNAGNREYLN 1140  
 QY 1196 TPQGAAPQPPPPAPSPAFDNLVYWDQ-----DP-----PERGAPSPSTFKGTP 1239  
 DB 1141 TAQ-----PTCLSSGFSNPALMIGKSHQMSLDNPDYQDFPKETKENGIFKG-P 1190  
 QY 1240 TAENPEYGLDVP 1252  
 DB 1191 TAENPEYLRVAP 1203  
 RESULT 4  
 Q9YH40 PRELIMINARY; PRT; 1165 AA.  
 ID Q9YH40  
 AC Q9YH40;  
 DT 01-MAY-1999 (TEMBLrel. 10, Created)  
 DT 01-OCT-2000 (TEMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)  
 DE Receptor tyrosine kinase proto-oncogene.  
 GN XKRK.  
 OS Xiphophorus xiphidium.  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
 NX NCBI\_TaxId=8086;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RX MEDLINE=98241172; PubMed=9582016;  
 RA Dmitriyevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,  
 RA Altschmidt J., Scharf M.;  
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by  
 overexpression and mutational alterations.";  
 RL Oncogene 16:1681-1690(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RA Scharf M.;  
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; U53471; AAD10500.2; -;  
 DR HSP; F11362; IFGK.  
 DR InterPro; IPR000345; Cyt\_c\_heme\_bind.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00669; pkinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; Fy; 3.  
 DR SMART; SM00219; TyrKc; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN\_1.  
 DR PROSITE; PS00343; GRAM\_POS\_ANCHORING; UNKNOWN\_1.

DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS50011; PROTEIN KINASE DOM; 2.  
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.  
 KW ATP-binding; Kinase; transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 1165 AA; 129614 MW; 7f7EB38D8771A74E CRC64;

Query Match 38.8%; Score 2647; DB 13; Length 1165;  
 Best Local Similarity 44.6%; Pred. No. 1,7e-191;  
 Matches 571; Conservative 164; Mismatches 388; Indels 156; Gaps 32;

QY 1 MELALGRMGILLALLPPG-AAST---CYCTGDMKLRPASBETHLMDLRLLYQCCQY 55  
 DB 4 LLEL-----LLELLLSIGRCSTDPDRKVCQGTSSNQMTM---LDNHYLLXKKKMYGSCNV 56  
 QY 56 VQGNLELTPTNLSFLDIOEVQGVLIANNOVQVPLQRLIRYRGTQLFEDYVALA 115  
 DB 57 VLENLEITITQENDDLSFLOSIGQGVYLIANNEVSTIPLVNLRLRGONTLEGNFTLL 116  
 QY 116 VLDNGDPLNNTPTVTGASFGGLRELQSLSTELIKGVLIQNPQLCYQDTILMKDIFKH 175  
 DB 117 VMSNYOK-NPSSP--DYVQVGLKQLSLNTEILSGVKNVSHNPLLCNVEITIMWDIVDK 173  
 QY 176 NNQALTLITDNRBRACHPCSPMKCSRCWGESESDCQSLTRTVCAAGC-ARCKPLPTD 234  
 DB 174 TSNPTMNLIRHAFERQCKDPCGVNASCNAPRGHCQKTKLCAEQCNRRRGKRPID 233  
 QY 235 CCEHQAAGCTGPRHSDCLALHNHSGICELHCPALVTYNTDTFESMPNDEGRYTFGAS 294  
 DB 234 CCEHQAAGCTGPRATDCLACRPDNDGCTCKTPPKIXDIVSHQVNDPNIKYTFGAA 293  
 QY 295 CVTACPRNYLSTDVSGCTVCPHNCVETAEDEGQREKSKCAVCYGLM----QYI 350  
 DB 294 CVECEPSNYVYVTE-GACVRSCKAGMLEVD-ENGRRSKPDDGVCPRKDDIGISLNTI 351  
 QY 351 KANSKFIGITELEFAGCKKIFGSLAPLPESFDGDPASNTAPLOEQLQVETLEITGYL 410  
 DB 352 ANVSTNG-----SESNCTKINGDIILNRNSFEDGPHKIGPMPENHMLNTYKELTGYI 407  
 QY 411 YISAMPDSDLDSVFQWLYIRIRILHNGAYS-LTLOGLSIMLGLRSLRELSSGLALH 469  
 DB 408 VIMWPEPNTSLSFQWLEIRRTTFSGRFSFVYVSHLQWGLSLKEVASAGVILK 467  
 QY 470 HNTHLCEVHTVPMDOLEFRNPHQALHNTANRPEDECEVGEJACQQLCARGCMGPRTOCV 529  
 DB 468 NTPQLRVASTINRRRLPESSEDQILEVDART-----ENQTCNNEGSEDCMGPGTMCV 520  
 QY 530 NCSQFLRGCEVEBCRYLQGLPREYVNAHCLPCHPECCPFQNGSVTCFGEADQCYACAH 589  
 DB 521 SCHLVDRGRCVASCMLLOGEPREAOVDGRCVQCHOECIVQTSILTCYGGPANCSKCAH 580  
 QY 590 YKDPFCVAPCPGSKVDLSYMPIMKPEDEGACOPCPINCTHSCVDLDDKGPAPQRAS 649  
 DB 581 FQDQPCQICPRCPHMLGSDPTL-TWKADKMGCCQCHONCTGCCGPGISGGRCD-IYS 638  
 QY 650 PLTISVAVVGLILVVLGVVFIL-----IFNFTVSEFWLRVVKYASHLBELTPSGAM 704  
 DB 639 HSSLAVGLVSGLLITVALIIVLLRRRIKRIKRIIRLLQCKE---VEPLTSGQA 694  
 QY 705 PNOAQRILKETELRKRYKLGSAGFGVYVGIWIPGGEVNIKIPVALIKYLRNTSPKANE 764  
 DB 695 PNOAFRIKETEIRKQDRVLSGAFGVYVGLNPGENIRIPVALIKYLRNTSPKANE 754  
 QY 765 ILDEAVYMGVGVSRLLIGLCTSTVQLVLTQMLPEYGLLDHRENRRLSGQDLNLC 824  
 DB 755 VLDEAVYMGVGVSRLLIGLCTSTVQLVLTQMLPEYGLLDHRENRRLSGQDLNLC 814  
 QY 825 MQIAKMSYLEDVRLVHDLAANVLYKSPNHKIDFGCLARLLIDELIYHAAGKVPY 884  
 DB 815 VQIAKMNLYEEHVLVHDLAANVLYKSPNHKIDFGSKLLTDEKEYOAGHGKVPY 874  
 QY 885 KMALESILRRFTHSDVMSYGVYVMEIATFGAKPYDGIAPAEIPDLLEKGRPLQPPY 944  
 DB 875 KMALESILQWYVTHSDVMSYGVYVMEIATFGSKPYDGIAPAEIPDLLEKGRPLQPPY 934

QY 945 CTIDVYMIWKCMIDSECRPRFRELVSFSMARDPQRFVYIYONEDLGPASPLDSTFYR 1004  
 DB 935 CTIEVYMIWKCMIDSPSRPRFRELVSFSMARDPKSYVLYIQ--NLSPSPDRLLS 991  
 QY 1005 SLEDDMDKDLVDAEYVLPQGFCCPPAPAGAGVWHRRSSSTRSGGDLTGLSPS 1064  
 DB 992 RLSSSD--DVVDADEYLL-----RYKRIN-RQSS----- 1018  
 QY 1065 EEEARSPPLAPEGAGSDVFDGDLGKAQKLSLPTHDPSFLQRYSEDPY-PLPSETD 1123  
 DB 1019 -----EPCIPNGH-----PVRENSIALRYISDPTQMALEKLD 1052  
 QY 1124 GYVAPLTCSPQPEYVNOQDVRPQP-----PSPRE-----GPLP-AAAPGATLERAKT 1170  
 DB 1053 GH-----EYVNOQPSSETSSRLSDIYNVNYEDLIDGMQPVLSLSEALTNFSRPEY 1102  
 QY 1171 LSPGKGVKVDYFAFGAVENDEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPERGA 1230  
 DB 1103 LNTNONSLS--PLVSSGSVDDEY--QAG-----YQAAF-----LPQTGA 1137  
 QY 1231 PSTFKGPTAEENPEYLG 1249  
 DB 1138 LTGNGMFLPAENLEHYGL 1156

## RESULT 5

Q9W6F6 PRELIMINARY; PRT; 1137 AA.

AC Q9W6F6; 01-NOV-1999 (TRENBLREL. 12, Created)  
 DT 01-NOV-1999 (TRENBLREL. 12, Last sequence update)  
 DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)  
 DE Receptor tyrosine kinase (Fragment).  
 GN ERB84.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 RN NCBI\_TaxID=9031;  
 RX MEDLINE=99263203; PubMed=10328884;  
 RA Dixon M., Lumsden A.;  
 RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in  
 RT embryonic chick hindbrain";  
 RT Mol. Cell. Neurosci. 13:237-258 (1999).  
 DR EMBL; AF121963; AAD31764.1; --  
 DR HSSP; P11362; 1FGK.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001368; TNR\_C6.  
 DR InterPro; IPR001245; Tyr\_Pkinase.  
 DR InterPro; IPR004019; YLP\_motif.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF01030; Recep\_L domain; 1.  
 DR Pfam; PF02157; YLP; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SMO0261; Fu; 3.  
 DR SMART; SMO0219; TyKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 DR PROSITE; PS00652; TNR\_NGFR\_1; UNKNOWN\_1.  
 KW Kinase; Tyrosine-protein kinase.  
 FT NON TER 1  
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F87DC84F CRC64;

Query Match

38.7%; Score 2636.5; DB 13; Length 1137;

Best Local Similarity 45.9%; Pred. No. 1e-190;  
Matches 525; Conservative 172; Mismatches 357; Indels 91; Gaps 26;

QY 161 LCYODTLMKDIFKNNQALATLIDTRSPACHPCSPCKSRCSSESSDCSLTFTVC 220  
Db 3 LCFADTLHMQDIVANPNASNTLIVPTNGSSCGCRCHKSCGT-RCMGPTENHCQTLTVC 61

QY 221 AGGC-ARCKGRLPDDCCHEQCACTGPKASDCLACIHFHNSGICELHGCALVTYNDTF 279  
Db 62 AECCDGRCYGYVSDCCHECAGCGSKPDTDFACMKNFNDSCGCTQCQTFVYNTTF 121

QY 280 ESMENPEGRYFGASCYACPYNYLSTDVGSCTVCPJLHNOEVAEDGTORCEKSKPCA 339  
Db 122 QLEHNNAKTYTGAFCYKCKPHNV-VDSSECVACASSKMEV-FEENIGICKRPTCICP 139

QY 340 RVCYGLGMOYIKANSKFIQTELE-FAGCKKIFSGSLAFEPSPGDASTAPLPQEQLO 358  
Db 180 KACDGIQGSU-VAQOTVDSNIDKFNCTKINNLFLVTGIGHDEYHTIAINPEKN 238

QY 399 VETLEETITGLYISAMPDSDLPDVSFONLOVIRGLIHNGVSLTQGLISWLGRL 458  
Db 239 IFOTVREITGLINQSPENMTDRVSNLVTIGRALYSGSLILKQGITSLQFQSL 298

QY 459 RELSGALIHNNHLCFVATVPWQLEFRNPHQALLHTANPEDECVSEGLACHQLCARG 518  
Db 299 KQISAGNIYITDNNLCTYHTVNTSLFSTPSQKTVIHRNKAENCTADGVCNELCSD 358

QY 519 HCMGPPTOCVNGOPFRCGEVCECRVLOGLPREYVNAHCLPCHPECOP-ONGSVTFC 577  
Db 359 CGWGPPOCLSKRFIRGTCIESCNLYDEFEFPANGSCVCMCDQCEMENNMTTCY 418

QY 578 GPEADQCVACAHYDPPFCVACPSGVKPDLSYMPIMKPPDEGACOPCPINCTHSCVDL 637  
Db 419 GPGDHCTKCHFPDGPVCEKCPDGLQGANF--IFKYADEDECHCHPCHNCTQGRGP 476

QY 638 DDGCG-----PAEGRASPLTISVAVV-GILLVYVLGVNPGILLFNNTVSPWL 685  
Db 477 ASHDCIYFWTRQSTLPQHAR-TPL-IAAGVIGLFTIVMGLTFAYVNRK----- 526

QY 686 RVPKVSASH-----LELTPSGAMPQAOIRLKELELARKVYLGSGAFVYKGIWI 738  
Db 527 SIKKBRALRRELETELVELVPLTPSGAPQAOIRLKELELARKVYLGSGAFVYKGIWI 586

QY 739 PDGENVKIPIVAKVIRENTSPKANKELIDEXVYVAVGVSPIVSRLLGICLTSTVQVLTOL 798  
Db 587 PEGELVKKIPIVAKIINETTPKANEFEDEALIMASMDPHLVLLGVCLSPITQVLTOL 646

QY 799 MPYGLDHEVRENRGLSGODLLNMCQIAKMSYLEDVRLVHEDLAARNVLVKSPNHVK 858  
Db 647 MPHGLLDVYHKNKNIQSOLLNMCQIAKMSYLEDVRLVHEDLAARNVLVKSPNHVK 706

QY 859 ITDFGLALLDIDETEHYADGGKVPDKMMALBSILRRRFTHQSDVSYGVTVWELMTFGA 918  
Db 707 ITDFGLALLDIDETEHYADGGKVPDKMMALBSILRRRFTHQSDVSYGVTVWELMTFGA 766

QY 919 KPYDIPAREIPDLLEKEGRLLPQPICTIDVYIMVCMIDSECRPFRELVSEFSMA 978  
Db 767 KPYDIPAREIPDLLEKEGRLLPQPICTIDVYIMVCMIDSECRPFRELVSEFSMA 826

QY 979 RDQRFVYIIONED-LGPASPLDSTFYRSLLEDDMDGLVDAEYVLVPOQGFCCDPAPGA 1037  
Db 827 RDQRFVYIIONED-LGPASPLDSTFYRSLLEDDMDGLVDAEYVLVPOQGFCCDPAPGA 885

QY 1038 GGMVHHRSSSTRSGGDDTLGLEPSESEAPRS-PLAP-SEGSGSVFPGDGLGMAK 1094  
Db 886 RTRIDSNKQVYRGGYAAEGV-PMYRAPGCIIPAPVAGATATIFEDTCNGTLR 944

QY 1095 GLQSLPDPSPLOQYSEDPIVPLPS-----ETDGYVAPLTCSPQDEYVNOQDPVPOP 1147  
Db 945 KOVATLAKEDSSTQYSDAPTVIFERVIRGELDEDDGYMTPMRDKPTDYNPEVNFV 1004

QY 1148 PSPREGPLPAA-RPAGATLERAKTLSPKNGVYKVF-----AFGAVENPEYLTLPQ 1198

Db 1005 SRKXGDDLOAVDNPEYHN-----APNGPAADEYVNEPLYNTFANTLENAEYL--- 1054

QY 1199 GGAAPQHPPEPAPFSAFDNLVYDODPPERGA--PSPFKGTPT-----AE 1242

Db 1055 -----KNLPERAKKAFNDPDMNSLPPRSLTQHPDYLOEYSTFYKONGRIRPIVAE 1109

QY 1243 NPEYL 1247

Db 1110 NPEYL 1114

RESULT 6  
P79754 PRELIMINARY; PRT: 1328 AA.

ID P79754  
AC P79754  
DT 01-MAY-1997 (TEMBREL. 03, Created)  
DT 01-MAY-1997 (TEMBREL. 03, Last sequence update)  
DT 01-JUN-2002 (TEMBREL. 21, Last annotation update)  
DE Erib3.  
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;  
OC Tetraodontidae; Takifugu.  
OX NCBI\_TaxID=31033;  
RN (1) SEQUENCE FROM N.A.  
RP MEDLINE=9917347; PubMed=1007531;  
RA Gallner K., Brenner S.  
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu  
rubripes".  
RL Genome Res. 9:251-258(1999).  
DR EMBL; AF056166; AAC34391.1; -  
DR HSSP; P11362; IFGK.  
DR InterPro; IPR000494; EGFR\_L domain.  
DR InterPro; IPR002174; Euk\_Pkinase.  
DR InterPro; IPR001245; Tyr\_Pkinase.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF00069; Pkinase; 1.  
DR Pfam; PF01030; Receptor\_L domain; 2.  
DR Prodom; PD000001; Euk\_Pkinase; 1.  
DR SMART; SM00261; FUR\_3.  
DR SMART; SM00219; Tyrc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
KW ATP-binding; Transferase.  
SQ SEQUENCE 1328 AA; 148613 MW; A33039258B647E9 CRC64;

Query Match 33.2%; Score 2265; DB 13; Length 1328;  
Best Local Similarity 39.8%; Pred. No. 1.8e-162;  
Matches 513; Conservative 160; Mismatches 418; Indels 198; Gaps 33;

QY 9 WGLLLALLP--GAASDQ---VCTGDMKRLRPASPTHLDLRLHYOGCQVVOGNLEL 62  
Db 4 WRLILNCVASRLAAASQGEAVCPETQNGLSGQENQVNLKORYKCEIIMGNLEI 63

QY 63 TYPTNASLFLDIDIEVQGYLIANNQRYVQLRLRYRGQLFEDNYAALVNDGCP 122  
Db 64 TQTESWDFEFLLTIREYGYVLIAMNHFOELPLGQLRIRGSLYERFASVFLN--- 120

QY 123 LNNTPVYTGSPGLRELOLRSLTEILKGVLIQRPOLCYODTILMKDIFKNNQALAT 182  
Db 121 ----YRDS--PSGLNQLGIMNLTLELDGVOGINNKYLRYGMYVYWRIL--RNMDAPLE 173

QY 183 LIDTNRSPACHPCSPMCKSRCSSESSDCQSLTRVYACAGC-ARCKGRLPDDCCHEQCA 241  
Db 174 IQNGERGVCH--KSC-GNYCWGPKDQCOILTKTVCAPOCNDRCFGSPRCCHIDEA 229

QY 242 AGCTGPHSDCLACIHFHNSGICELCPALVTYNTDTFESMPNPEGRYTFGASCYACPY 301  
Db 230 AGCKGRLPDDCCHEQCACTGPKASDCLACIHFHNSGICELCPALVTYNTDTFESMPNPEGRYTFGASCYACPY 301

```

QY 302 NYLSTVSGCTLVCLPLHNOEYTAEDGTQR--CEKSKRCPACVAYCYGLAQWYIKANSKFIGIT 360
D 290 HFV-VDSGSCVSVPPPKMEV--ERGSQROCELCSGICPVYCESTGE---GRQYDSS 342
QY 361 ELE-FACCKKIFGSLAFPEFSDGDPASNTAPLOPEQLQVFTLEITGYLYISAMPDSL 419
D 343 NIDSFINCTKIQGSLHFLVGLIGDDDEKPNVPLDAKLEVFRTREITDILNTIQSWPKEL 402
QY 420 PDLISFQNLQVIRRIIHNAGAVSLTLOGLSWGLNLSBELSGSLALIHNNHLCVHT 479
D 403 NDLSVFSSLLTTIQGRSLFKRPSLWMWRIPTLISGLSLBELISDGSVYISQNAHLCTHHT 462
QY 480 VPMDOLEFRNPH-QALLHTANRPEDECVEGELACHOLCARHCWMPGPPTQCVNSQPLRG 538
D 463 VMTQLEFRGSVRANSINSRPMACVADGRVCDPLCSDSGCMGPGDQLCSRNVGRHG 522
QY 539 ECVBECKVLOGLPREYVNAH-CLPCHPECOPOKNGSTTCGPREDQCVACAHKXDPFCV 597
D 523 TVAGCHFNNGIPREFAGLNGVCAVCHPECKPOTGKASCTGPGADBEOMACTKFRDGYCM 582
QY 598 ARCPGVPDLISYPIWKPFDEGACOPCPINCTHSCVDLDDKCPAEOAPSLTISVA 657
D 583 SSCPAGVN-DEGKLLFKFVRBECHGCPHQNCTQSGSGPLNDC---LEAHLTLTSSG 638
QY 658 VVGLLVVVLGVGVGLIFNNFTYSFWLRVPRKV-----SASHLEPLTPSGAMPN 706
D 639 ITGIALGVPAGLICLVLF--FLQMLYHRGLAIRKRAMRYLESSEGFEPPLP-GEKGT 695
QY 707 QAKRILKETELRKVKVYLSGAFSTYKGIWIPGDEWVKIPVAIKVIRENTSPKANEIL 766
D 696 KVAHRIKPSDLRIKIKPLGSGVFSTVSKGFWIPGEVYKIPVAIKTIQDSSGQTFEIT 755
QY 767 DEAYVAVGSPVYSRLLIGCLISTVOLVTLQMLBYGCLLDHVENRGLSGODLLMCMQ 826
D 756 DHLTMSGLDHPYIVRLIGICPGCTCLQVLSHSGSLBLHRIHQKSLDPPQRLMVCQ 815
QY 827 IAKMSYLEDVRLVRLHRLAARNVLYKSPNHVKITPDLALDLDEHYHADGKXVITK 886
D 816 IAKMYLLEHVRVVKLARNILKNDYQVQISDGVADLVLPDDKKVYSETKTFITK 875
QY 887 MALESILRRRTHOSDVSYGVTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPIC 946
D 876 MALESILFRRTHOSDVSYGVTVMELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPIC 935
QY 947 IDVYIMVKKCMIDSECRPRRELVSFSSRNADPQRFVITQNEIDLSPASPLDSTFYS 1006
D 936 IDVYIMVKKCMIDSECRPRRELVSFSSRNADPQRFVITQNEIDLSPASPLDSTFYS 980
QY 1007 LEDDDMGDLVDAEYVLPQOGFCPCDPAFGAGVWHRHRSSTRSGGDLTLGLEPSEE 1066
D 981 EDGSGMGEPL-----RGSER--GLLEADLEDEE 1007
QY 1067 EAPRSPLAPSEAGSDVFDGDLGNG--AAKGLSLTHDSPLQ-----RYSED 1113
D 1008 E-----GLGDRFATPSLOPSPSWSTSPQINSYMMTQLRYD-- 1044
QY 1114 PTVLEPSETDGVAVPLTCSPOP-EYVNO-----PVRPQPSREGPL--P 1156
D 1045 ----FAVSGGHIGIYLMSPBPVDTIRQLMWQRSLSSVATLPRKSAFRSSSEALCED 1100
QY 1157 AARPAGATLEBAKTLSPGKGVKDVAFAGAVENPYYLTPQGAADQPHPPAFSPAFD 1216
D 1101 GAQCAGIFRVR-----FSGERN-----FOG-- 1122
QY 1217 NLVYMDOPPERGAPSTFKGTPTAENPE 1245
D 1123 ----QBRKSTASSPSFKTMADDEDE 1146

```

RESULT 7  
 Q9BIH9  
 ID Q9BIH9 PRELIMINARY; PRT; 1433 AA.  
 AC Q9BIH9;

```

DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SU4;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor";
RL Submitted (NOV-2000) to the EMBL/Genbank/DBD databases.
DR EMBL; AJ301655; CAC35008.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000345; CycC_heme_bind.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser_th_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; Fv; 7.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYKc; 1.
DR PROSITE; PS00190; CYTOCHROME_C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
KM receptor.
FT NON_TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 29.5%; Score 2012.5; DB 5; Length 1433;
Best Local Similarity 32.3%; Pred. No. 2.8e-143;
Matches 468; Conservative 202; Mismatches 390; Indels 387; Gaps 38;

QY 26 CTGTDMKRLRPASPEHLDMLRHLVYQGCQVQGNLELYLFTFNASLSFLQDIOEVGYVL 85
D 1 CIGTNGRMVSPANREHYKRLDRYNTCTYVGNLEITWQNTIDLNFQHIRVGYVL 60
QY 86 IAHQVROYPLRLRIRVGTOLF-----EDNYALAVLDNGDPLNNTTPVTGASPGSLREL 140
D 61 ISYDLPQVILPRDLIRKRTTFKLNKEEAYGLV-----SFHMTL 104
QY 141 QURSLTEILKGVLIQORNPOLCYODTILMKDI-FHKNOALATLIDTNRSACHPSPKC 199
D 105 ELPALRDILGSGVGFNNYNLCHMSINWEEILAPQTSMTQTFNPSSEBVCPCHPSC 164
QY 200 KSGRCWGESSEDDCOGSTRVCAAGCA--RCKRPLPDCCHBECAAGCTPRKSDCLACH 257
D 165 EVG-CWEGEAGHNCQRFSLKNSPQCSQGFPRKPRECCNHLRCAGCGCTGPTSDCLACK 223
QY 258 FNHSGICEHLHCAALVYNTDTFESMPNPEGRTTFGASCVTACRYVVLSTDVSGCTLVCP 317
D 224 FYDDGCKQECRPMQIYNPTNFWMEPNPGKAYATGCRKP-EHLKNDACVYRKCKX 282
QY 318 HNOBTAELEGTORCEKCKPCARVYCYGLAQWYIKANSKFIGITLEFAGCKKIFGSLAF 377
D 283 GMPQNSE-----CVPCKGVCPKTCPEGDI-----VHSDNIG-----NYXDCITIEGSL 329
QY 378 PSFPGDPAANT-----APLOEQLQVFTLEITGYLYISAMPDSLPLISVQNLQ 429
D 330 DGSFDFQGVYNTNFSFGPRYIKIDPDLRELFVSTVAEITGFINQAHHRFTLTINFRMLE 389

```

430 VIRGRIANGAY-SLTLQGLGIGSWLGRSLRELGSGLALIHNTLHCFTVTPWDLFRN 468  
 390 VVGRGRLKEMLFASVYIVKTSILKSLKSLKRVNVSGLVILENSDLCEVEDIMSEIKKS 449  
 489 PHOALLTANPRECECEGEGJACHQOLARHGCHWPGTTCVNSQFLRGECEVECEVLO 548  
 450 SDHEWVQKNNATECHEEGMECEQCSKAGCKWKGPEQCECNVKKKCKLDSCK--- 506  
 549 GLPREY-VNARHCLPCHPECCOPONGSVTCFGEADQCVACAHYKDPFCVARGP----- 601  
 507 SLRPLYSVDSKTCGDCHQCEKD-----FCGPNEDNGSCNNVXKGRFCVABEPTTGAM 561  
 602 -----SGVKPDLSTMPKTPD----- 618  
 562 NGTCINCHKTGCGRPRDTIAPDGCISCDKAIIGSADAKIERCLMKDESCPDGYSDYVL 621  
 619 -EEG----- 621  
 622 QEEGRLKLSGKAVCRKCHPRCKKCTGTFHGFQCFQCTGYKKGECEDECPDFYANEE 681  
 622 -ACQPCPINCT-----HSCVDL-----DD-----KCGPABG----- 646  
 682 TRICLPCHQECRGCHGLDDHHEERNLKLPEGDYDVAATPTTCVSNCPASHPKRFPQEA 741  
 647 -----RASPLTSIVSAVVGILLVVLGVVFGI--LIEN--NFTVSEWL 685  
 742 GRTGFCYCADSMOSGLRIEFQTKYKMGVMAILLVCVGFIAFVLFRRKRRKDAVNM 801  
 666 RVPKVASHLEPLTPSGAMPNQAOMRIKETELRKVYLSGAGFQYKGIWIDGENVK 745  
 802 TMALAGCEDESEPLRPSNVGPNLTKRLIKAEIRRGVLAGAGAGRVFKGVWMEGESYK 861  
 746 IPAIKVLRENTSPRANKELLIDEAYVMAGVSFVYSLIGCLSTSTQVLQVLMPCYCL 805  
 862 IPAIKVLEWMSGSSKEPLBEATIMASVEHPNLIKLLAYCMISQKMLLIQLPGLCL 921  
 866 DLYRNKCKIGSKALINWSTQJARMAYLEERLVRDLAARNVLVQCPSCVKTIVFGLA 981  
 866 RLDDIDETEHADGKVPKMALESILARRFTIOSDVMSGVVWELMTGAKPYGIP 925  
 982 KLLPDSSEYRAAGKMPKIMWALECIRHRAVFTSKSDWAGITWELLTYGAPYEIVP 1041  
 926 AREIPDLLEKGERLPOPEICTIDVYIMVWCMMDISECRPRFRELVSFEMADPQRFV 985  
 1042 AKDVELIEIGHKLPQDPCISLDVYCLISCVLADADARPTFOAETPAKAKADPGRYL 1101  
 986 VIQNEDDLGPASFLDSTFYRSLLEDMDKGLV----- 1016  
 1102 MI-----PGDKFMRLLPSYTNODEKDLIRTLAPVMAAAAAAAGASNVDPSTIA 1152  
 1017 DAEYLVFPOQGFCDPAPAGAGWVHHRRSSSTRSGGDDLTLGLEPSEEAAPRS----- 1071  
 1153 EIDEVLYQPKTRISVILPGBSA-----VEPS-DEMPSLRCK 1188  
 1072 -PLAF--SEGAGSVFPODGLMGAAKGLQSLPHDPBPLORYSEDPVLPBPSETGVYA 1127  
 1189 DPLKPDDETDGHEKVEV-----GVGGIR-----LNLPLDEDDYLM 1222  
 1128 PLTCSPOPEYVQPVRRPQPSFREGPLPAARPAATLERAKTISFGKNGVAVKVFARFG 1187  
 1223 P-TCOSQ-----NOS-----TPG-----YMDLIGIPA 1243  
 1188 AVENPEYV-----TPQGAAPQHPHPAPAFSPAFNLYWQDPPERGAPSTFKGT 1238  
 1244 SYDNBEYLMGSTQAIAGLAGSGMG--PHTPP-----PMTPTGGM 1280  
 1239 PTAENPE 1245  
 1281 PTHQHSQ 1287

RESULT 8  
 Q9UK79 PRELIMINARY; PRT; 419 AA.  
 AC Q9UK79;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Heretatin.  
 GN HER-2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=99415951; PubMed=10485918;  
 RA Doherty J.K., Bond C., Jardim A., Adelstein J.P., Clinton G.M.;  
 RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted  
 RT autoinhibitor";  
 RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Doherty J.K., Clinton G.M., Adelstein J.P., Evans A.J., Henner W.D.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF177761; AAD5609.2;  
 DR InterPro: IPR000494; EGFR\_L domain.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L domain; 1.  
 DR SMART: SM00261; Fv; 1.  
 SQ SEQUENCE 419 AA; 45472 MW; FEC1BE347E2D030C CRC64;  
 Query Match 27.5%; Score 1871; DB 4; Length 419;  
 Best Local Similarity 98.8%; Pred. No. 2, 4e-133;  
 Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 1 MELALCRWGLLALLPGAASTOVCTGDMKRLPASBETHLDMRLHYOGQVQVNM 60  
 DB 1 MELALCRWGLLALLPGAASTOVCTGDMKRLPASBETHLDMRLHYOGQVQVNM 60  
 QY 61 ELYTLPNASTSLDIOIQEYGVVLAHNOVOPVORLRIYRGTOLEFEDNVALVDNG 120  
 DB 61 ELYTLPNASTSLDIOIQEYGVVLAHNOVOPVORLRIYRGTOLEFEDNVALVDNG 120  
 QY 121 DPLNTPVYVGSAGPGIRLQSLSTELIKGVLIOGNPOLCYQDITLWKDIFKXNOQA 180  
 DB 121 DPLNTPVYVGSAGPGIRLQSLSTELIKGVLIOGNPOLCYQDITLWKDIFKXNOQA 180  
 QY 181 LTLIDNRSRACHPCSPMCKSGRCMESESDCSLRTYVAGGACACKGLPTDCHEQC 240  
 DB 181 LTLIDNRSRACHPCSPMCKSGRCMESESDCSLRTYVAGGACACKGLPTDCHEQC 240  
 QY 241 AAGCTGKSHDCLACHFNHSGICELHCPALVYNTDFTESMPNPGRYVFGASCYTACP 300  
 DB 241 AAGCTGKSHDCLACHFNHSGICELHCPALVYNTDFTESMPNPGRYVFGASCYTACP 300  
 QY 301 YNLTSDVGSCTLVCPILHNOEVTAEQTQRCCKSPCARVCYGL 345  
 DB 301 YNLTSDVGSCTLVCPILHNOEVTAEQTQRCCKSPCARVCYGL 345  
 RESULT 9  
 Q8R2X1 PRELIMINARY; PRT; 367 AA.  
 ID Q8R2X1;  
 AC Q8R2X1;  
 DT 01-JUN-2002 (TREMblrel. 21, Created)  
 DT 01-JUN-2002 (TREMblrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Hypothetical 40.2 kDa protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;



RN (1)  
 RP SEQUENCE FROM N.A.  
 RA Strausberg R.  
 RL Submitted (Apr-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC027080; AA027080.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;  
 Query Match 25.5%; Score 1739; DB 11; Length 367;  
 Best Local Similarity 88.0%; Pred. No. 2e-123;  
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;  
 QY 887 MALESLRRRFTHQSDVMSGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 946  
 DB 1 MALESLRRRFTHQSDVMSGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 60  
 QY 947 IDVYIMVKCMIDSECRPRFRELVSFSSRMARDPQRFVYIQLNEDLGASPLDSTYRSL 1006  
 DB 61 IDVYIMVKCMIDSECRPRFRELVSFSSRMARDPQRFVYIQLNEDLGASPLDSTYRSL 120  
 QY 1007 LEDDDMGDLVAEEYLVPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 1066  
 DB 121 LEDDDMGDLVAEEYLVPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 180  
 QY 1067 EAPRSPLAPSEAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYV 1126  
 DB 181 EAPRSPLAPSEAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYV 240  
 QY 1127 APLTSPQPEYVNOQDVPRPQSPREGPLPAARAGATLERAKTILSPGKGVKDVFAFG 1186  
 DB 241 APLTSPQPEYVNOQDVPRPQSPREGPLPAARAGATLERAKTILSPGKGVKDVFAFG 300  
 QY 1187 GAVENBEYLTPOGGAAPQPPPPAFSPAFNLVYNODPERCAPSTKGTPTANPEY 1246  
 DB 301 GAVENBEYLTPOGGAAPQPPPPAFSPAFNLVYNODPERCAPSTKGTPTANPEY 360  
 QY 1247 LGADVPEY 1253  
 DB 361 LGADVPEY 367  
 RESULT 10  
 Q8WYVO PRELIMINARY; PRT; 412 AA.  
 AC 08WYVO:  
 DT 01-MAR-2002 (TREMBlrel. 20, Created)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Hypothetical 44.7 kDa protein.  
 GN PP3659.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,  
 RA Wan D.F., Gu J.R.;  
 RT "Novel human cDNA clones with function of inhibiting cancer cell  
 RT growth."  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF18349; ALU5856.1; -  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR InterPro: IPR004019; YLP motif.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00219; TyrKc; 1.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Hypothetical protein.

SQ SEQUENCE 412 AA; 44702 MW; 034397F3F27D2BC CRC64;  
 Query Match 24.9%; Score 1697.5; DB 4; Length 412;  
 Best Local Similarity 80.5%; Pred. No. 3.3e-120;  
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;  
 QY 887 MALESLRRRFTHQSDVMSGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 946  
 DB 1 MALESLRRRFTHQSDVMSGYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPDPIC 60  
 QY 947 IDVYIMVKCMIDSECRPRFRELVSFSSRMARDPQRFVYIQLNEDLGASPLDSTYRSL 1006  
 DB 61 IDVYIMVKCMIDSECRPRFRELVSFSSRMARDPQRFVYIQLNEDLGASPLDSTYRSL 120  
 QY 1007 LEDDDMGDLVAEEYLVPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 1066  
 DB 121 LEDDDMGDLVAEEYLVPQGFPCPDPAAGAGVHHRRSSSTRSGGDLTLGLEPSEE 180  
 QY 1067 EAPRSPLAPSEAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYV 1126  
 DB 181 EAPRSPLAPSEAGSDVFDGDLGMAAGLQSLPTHDSPLQRYSEDPVLPSETDGYV 240  
 QY 1127 APLTSPQPEYVNOQDVPRPQSPREGPLPAARAGATLERAKTILSPGKGVKDVFAFG 1186  
 DB 241 APLTSPQPEYVNOQDVPRPQSPREGPLPAARAGATLERAKTILSPGKGVKDVFAFG 300  
 QY 1187 GAVENBEYLTPOGGAAPQPPPPAFSPAFNLVYNODPERCAPSTKGTPTANPEY 1246  
 DB 301 GAVENBEYLTPOGGAAPQPPPPAFSPAFNLVYNODPERCAPSTKGTPTANPEY 360  
 QY 1219 YWMD-QDPPER-----GAPSPFKGTPTAN 1243  
 DB 361 WMTQCEPQEGVRRSPDVSSGREGTLTSGIKRWEKGPPTSGTCHARN 410  
 RESULT 11  
 Q86712 PRELIMINARY; PRT; 729 AA.  
 AC 086712:  
 DT 01-NOV-1996 (TREMBlrel. 01, Created)  
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)  
 NT Polypeptide.  
 UN POLYPEPTIDE.  
 OS Avian fous-associated virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.  
 OC NCBI\_TaxID=11950;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94203659; PubMed=8152791;  
 RA Venstrom B., Raynocheck C., Jansson L., Doederlein G., Uhroak V.,  
 RA Johnson A., Beug H.;  
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
 RT different transforming capacities."  
 RL Oncogene 9:1307-1320(1994).  
 DR EMBL: S69372; AAC60725.1; -  
 DR HSP: P03322; 1A6S.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR004028; Retro\_M.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF02813; Retro\_M; 1.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00219; TyrKc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 729 AA; 80649 MW; 84D2F6914EF61D63 CRC64;  
 Query Match 24.1%; Score 1643; DB 15; Length 729;  
 Best Local Similarity 53.8%; Pred. No. 1e-115;

Matches 344; Conservative 71; Mismatches 124; Indels 100; Gaps 14;

QY 582 DQVACAHYKDPFCVACRCPGKPDLSYMPIMKFPDEGACOPCEINCTHSCVDLDDG 641  
 DB 152 DHCMKCAHFDGPHCVACRCPAGVLENDTL-VWKADANAVCQLCHPCTRGCKGGLG 210  
 QY 642 CPAEQRASPLTISVAV-GILLVVLGVVFGILFNNFTV-SFWLRVPKVSASHLEPLT 699  
 DB 211 CP---NGSKTPSIAGVVGGLCLVVGGLGLYLRHRHIVKRTLRLLQERLEVEPLT 267  
 QY 700 PSGAMPNQAQMRILKETELRKVKYLSGAGFTVYKGIIPDGENVKIPVAIKYLRNTSP 759  
 DB 268 PSGEAPNQAHLRIKETEFKKVKYLSGAGFTVYKGIIPDGENVKIPVAIKYLRNTSP 327  
 QY 760 KANKELIDEAAYVAVGSPVYSLGICITSTVQLTQMLPYGCLLDVHRENGRGLGSD 819  
 DB 328 KANKELIDEAAYVAVGSPVYSLGICITSTVQLTQMLPYGCLLDVHRENGRGLGSD 387  
 QY 820 LNNMCQIAKGMVSYLEDVRLVHRLAARNVLYKSPNHVKTIDFGARLLDIDETEHADG 879  
 DB 388 LNNMCQIAKGMVSYLEDVRLVHRLAARNVLYKSPNHVKTIDFGARLLDIDETEHADG 447  
 QY 880 GKVPIKMALESILRRFTHOSDVSYGTVWELMTFGAKPYGIPAREIPDLLEKGERL 939  
 DB 448 GKVPIKMALESILRRFTHOSDVSYGTVWELMTFGAKPYGIPAREIPDLLEKGERL 507  
 QY 940 POPICITIDVIMVIMKCMIDSECRPRELVSFSEPMARDPOSFVITQ-NEDLGPASPL 998  
 DB 508 POPICITIDVIMVIMKCMIDSECRPRELVSFSEPMARDPOSFVITQ-NEDLGPASPL 567  
 QY 999 DSTFYRLLEDMDGDLVDAEYLVPOQGFPCPDPAAGGVVHRRHSSSTRSGGDLT 1058  
 DB 568 DSTFYRLLEDMDGDLVDAEYLVPOQGFPCPDPAAGGVVHRRHSSSTRSGGDLT 603  
 QY 1059 LGLEPSEEARPSPL-----APSEAGSDVFPDGLGMAKGLQSLPTHPSPLQRYSED 1113  
 DB 604 -----SRTPLLSSLSATSNNSATNCID-----RNGGCHPVRDEDSFYQRYSSD 645  
 QY 1114 PTVPLEPSET-DGYVAPLTCSPQPEYVNOQDVRPQPSPREGRLPAARPAATLERAKTL 1171  
 DB 646 PTGNFLESIDGFL-----PAPEYVNO--LMPKPS-----T 676  
 QY 1172 SPKNGVVKDVF-----AFGAVENPEYL 1195  
 DB 677 AMVQNIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL 715

RESULT 12

ID 066714 PRELIMINARY; PRT; 567 AA.

AC 066714  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-1AR-2002 (TREMBLrel. 20, Last annotation update)  
 DE V-erbB protein (Fragment).  
 GN V-ERBB.  
 OS Avian rous-associated virus type 1.  
 CC Viruses; Retrovirdae; Retroviridae; Alpharetrovirus.  
 CX NCBI\_Taxid=11950;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94203659; PubMed=8152791;  
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
 RA Johnson A., Beug H.;  
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
 RT different transforming capacities";  
 RL Oncogene 9:1307-1320(1994).  
 DR EMBL: S69372; AAC60727.1; -.  
 DR HSSB: P11362; IFGK.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR001245; Tyr\_Kinase.  
 DR Pfam: PF00069; Pkinase; 1.

DR PRINTS; PRO0109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW Tyrosine-protein kinase.  
 FT NON\_TER  
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 24.1%; Score 1641; DB 15; Length 567;  
 Best Local Similarity 53.8%; Pred. No. 1e-115;  
 Matches 344; Conservative 71; Mismatches 124; Indels 100; Gaps 14;

QY 582 DQVACAHYKDPFCVACRCPGKPDLSYMPIMKFPDEGACOPCEINCTHSCVDLDDG 641  
 DB 3 DHCMKCAHFDGPHCVACRCPAGVLENDTL-VWKADANAVCQLCHPCTRGCKGGLG 61  
 QY 642 CPAEQRASPLTISVAV-GILLVVLGVVFGILFNNFTV-SFWLRVPKVSASHLEPLT 699  
 DB 62 CP---NGSKTPSIAGVVGGLCLVVGGLGLYLRHRHIVKRTLRLLQERLEVEPLT 118  
 QY 700 PSGAMPNQAQMRILKETELRKVKYLSGAGFTVYKGIIPDGENVKIPVAIKYLRNTSP 759  
 DB 119 PSGEAPNQAHLRIKETEFKKVKYLSGAGFTVYKGIIPDGENVKIPVAIKYLRNTSP 178  
 QY 760 KANKELIDEAAYVAVGSPVYSLGICITSTVQLTQMLPYGCLLDVHRENGRGLGSD 819  
 DB 179 KANKELIDEAAYVAVGSPVYSLGICITSTVQLTQMLPYGCLLDVHRENGRGLGSD 238  
 QY 820 LNNMCQIAKGMVSYLEDVRLVHRLAARNVLYKSPNHVKTIDFGARLLDIDETEHADG 879  
 DB 239 LNNMCQIAKGMVSYLEDVRLVHRLAARNVLYKSPNHVKTIDFGARLLDIDETEHADG 298  
 QY 880 GKVPIKMALESILRRFTHOSDVSYGTVWELMTFGAKPYGIPAREIPDLLEKGERL 939  
 DB 299 GKVPIKMALESILRRFTHOSDVSYGTVWELMTFGAKPYGIPAREIPDLLEKGERL 358  
 QY 940 POPICITIDVIMVIMKCMIDSECRPRELVSFSEPMARDPOSFVITQ-NEDLGPASPL 998  
 DB 359 POPICITIDVIMVIMKCMIDSECRPRELVSFSEPMARDPOSFVITQ-NEDLGPASPL 418  
 QY 999 DSTFYRLLEDMDGDLVDAEYLVPOQGFPCPDPAAGGVVHRRHSSSTRSGGDLT 1058  
 DB 419 DSTFYRLLEDMDGDLVDAEYLVPOQGFPCPDPAAGGVVHRRHSSSTRSGGDLT 454  
 QY 1059 LGLEPSEEARPSPL-----APSEAGSDVFPDGLGMAKGLQSLPTHPSPLQRYSED 1113  
 DB 455 -----SRTPLLSSLSATSNNSATNCID-----RNGGCHPVRDEDSFYQRYSSD 496  
 QY 1114 PTVPLEPSET-DGYVAPLTCSPQPEYVNOQDVRPQPSPREGRLPAARPAATLERAKTL 1171  
 DB 497 PTGNFLESIDGFL-----PAPEYVNO--LMPKPS-----T 527  
 QY 1172 SPKNGVVKDVF-----AFGAVENPEYL 1195  
 DB 528 AMVQNIYNNISLTAISKLPMSRYQNSHSTAVDNPEYL 566

RESULT 13

ID 064895 PRELIMINARY; PRT; 962 AA.

AC 064895  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Gag-V-erbA-V-erbB protein.  
 GN GAG-V-ERB-A-V-ERB-B.  
 OS Avian erythroblastosis virus.  
 CC Viruses; Retrovirdae; Retroviridae; Avian type C retroviruses.  
 CX NCBI\_Taxid=11861;  
 RN (1)  
 RP SEQUENCE FROM N.A.

RX MEDLINE=90206603; PubMed=1969616;  
 RA Bruskini A., Jackson J., Bishop J.M., McCauley D.J., Schatzman R.C.;  
 RT "Six amino acids from the retroviral gene gag greatly enhance the  
 RT transforming potential of the oncogene v-erb-B";  
 RL Oncogene 5:15-24 (1990).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 DR EMBL: X52209; CA36459.1; -.  
 DR EMBL: X52211; CA36459.1; JOINED.  
 DR HSSP: P10828; 2NLL.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR000536; Hormone\_rec\_1g.  
 DR InterPro: IPR001723; Strdhm\_receptor.  
 DR InterPro: IPR001245; Tyr\_Pkinase.  
 DR InterPro: IPR001628; Znf\_C4steroid.  
 DR Pfam: PF00104; hormone\_rec.1.  
 DR Pfam: PF00069; kinase.1.  
 DR Pfam: PF00105; Zf-C4.1.  
 DR PRINTS: PR00398; STRDHOMONER.  
 DR PRINTS: PR00047; STROIDFINGER.  
 DR ProDom: PD000001; Euk\_Pkinase.1.  
 DR ProDom: PD000035; Znf\_C4steroid.1.  
 DR SMART: SM00430; HOL1.1.  
 DR SMART: SM00219; TyKc.1.  
 DR SMART: SM00399; Znf\_C4.1.  
 DR PROSITE: PS00031; NUCLEAR\_RECEPTOR.1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP.1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM.1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR.1.  
 KM ATP-binding; DNA-binding; Nuclear protein; Receptor;  
 KM Transcription regulation; Transferase; Tyrosine-protein kinase;  
 KM Zinc-finger.  
 SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E495CE CRC64;  
 Query Match 23.1%; Score 1576.5; DB 15; Length 962;  
 Best Local Similarity 50.2%; Pred. No. 1,7e-110;  
 Matches 347; Conservative 71; Mismatches 154; Indels 119; Gaps 18;  
 QY 541 VEECRYVQGLPRE-YYNAR-HCLP-----CHPECO 568  
 DB 354 IEKQESYLLAFHYIYRGNIPHEWSKLMKRVADLMIGAHASRFLMKVECPTELS 413  
 QY 569 PONGSVTCFGEPAQOCVACAHYKDPFCVACRPSGVKPDLSYMPINKFPEBEGACQPCPI 628  
 DB 414 PGE-----VGP--DHCKKCAHFDGPHCVACCPAGVIGENDTL-VKRYADANAVCOLCHP 465  
 QY 629 NCHTSCVDLDDKCGPACQASPLTSIVSAVY-GILLVVYGVVFGILIFNNFTV-SFWLR 686  
 DB 466 NCTRGCKGPGLEGCP--NGSKTPSIAGVVGILLCLVVVGLGIGLYLRRIIRKRTLR 522  
 QY 687 VPRVSAHLEPLTPSGAMPNOQMRLIKETELRKVKVLSGAGCTYVKGWIPDGENVKI 746  
 DB 523 RLIGRELVEPLTPSGARPNOAHRLIKETEFKVKVGLGAGFTYVKGWIPDEGKVI 582  
 QY 747 PVAIKVLENTSPRANKELIDEAYVVAAGVSPYVSRLLGICLTSTVOLVQLMPYGCILD 806  
 DB 583 PVAIKELREATSPRANKELIDEAYVVAAGVSPYVSRLLGICLTSTVOLVQLMPYGCILD 642  
 QY 807 HVENRGRGLSGODLLNWCQIAGMSYLEVRLVHRDLAARNVLYKSPNNVKITDGLAR 866  
 DB 643 YIEHKDNGISQYLLNWCQIAGMSYLEVRLVHRDLAARNVLYKSPNNVKITDGLAR 702  
 QY 867 LLDIDETEVHADGKVPKIMMALESILRRFTHQSDVWSYGVVWELMTFGAPYDGI 926  
 DB 703 QLGADKEVHAEGKVPKIMMALESILRRFTHQSDVWSYGVVWELMTFGAPYDGI 762  
 QY 927 RELPDLLEKGERLPQPPICITDYYMIMVCKMIDSECRPFRELVSEFSMAADPORFV 986  
 DB 763 SEISSVLEKGERLPQPPICITDYYMIMVCKMIDSECRPFRELVSEFSMAADPORFV 822  
 QY 987 IQ-NEDLGPASPLDSTFYRSLLEDDDMGDLVDAEYLVLPQGGFCDDPAPGAGSMVHHH 1045  
 DB 823 IQDERKHLPSPTDSKFTYRLTMEEDMEDLVDAEYLVLPQGGFCDDPAPGAGSMVHHH 866

QY 1046 RSSSTRSGGGLTLLEPSEBEPAPPLAPSRGAGSDVPDGLMGAAKGLDPTHDS 1105  
 DB 867 NSPST-----SRTPLLSSLSATSN-----NSATKICIRNGCH--- 898  
 QY 1106 PLQRYSEDPTEVLPSEIDGYVAPLTCSPQPEYVNOPDVPQPPSPREGFLPAPAPAGAT- 1164  
 DB 899 -----PREDDGFL-----PAPEVVO--LMPKKPSTAMQNOYNYISLTA 937  
 QY 1165 LERATLSPKXGKGVKDVAFAGAVENPEYL 1195  
 DB 938 ISKLPMDSRYON-----SHSTAVDNPETL 961  
 RESULT 14  
 ID Q85468 PRELIMINARY; PRT; 545 AA.  
 AC Q85468;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE Avian Erythroblastosis virus (Tsj3) v-erbB gene.  
 OS Avian erythroblastosis virus.  
 OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.  
 OX NCBI\_Taxid=11861;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88217326; PubMed=2897102;  
 RA Scotting P., Venstrom B., Jansen M., Graf T., Beug H., Haymann M.T.;  
 RT "Common site of mutation in the erbB gene of avian erythroblastosis  
 RT virus mutants that are temperature sensitive for transformation";  
 RL Oncogene Res. 1:265-278 (1987).  
 DR EMBL: X06943; CA30024.1; -.  
 DR HSSP: P11362; 1FGK.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR001245; Tyr\_Pkinase.  
 DR Pfam: PF00069; kinase.1.  
 DR ProDom: PD000001; Euk\_Pkinase.1.  
 DR SMART: SM00219; TyKc.1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP.1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM.1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR.1.  
 KM ATP-binding; Transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 545 AA; 60899 MW; 140DCB8CA0FAPFA CRC64;  
 Query Match 23.0%; Score 1568; DB 15; Length 545;  
 Best Local Similarity 53.3%; Pred. No. 3.2e-110;  
 Matches 334; Conservative 68; Mismatches 133; Indels 92; Gaps 15;  
 QY 578 GPEADQCVACAHYKDPFCVACRPSGVKPDLSYMPINKFPEBEGACQPCPICTHSCVDL 637  
 DB 1 GP--DHCKKCAHFDGPHCVACCPAGVIGENDTL-VKRYADANAVCOLCHPCTGCKGP 57  
 QY 638 DDKGCPACQASPLTSIVSAVY-GILLVVYGVVFGILIFNNFTV-SFWLRPVKVSASHL 695  
 DB 58 GLEGCP--NGSKTPSIAGVVGILLCLVVVGLGIGLYLRRIIRKRTLRLOERELV 114  
 QY 696 EPLTPSGAMPNOQMRLIKETELRKVKVLSGAGFTYVKGWIPDGENVKIPVAIKVRE 755  
 DB 115 EPLTPSGARPNOAHRLIKETEFKVKVGLGAGFTYVKGWIPDEGKVIIPVAIKERE 174  
 QY 756 NTPSPRANKELIDEAYVVAAGVSPYVSRLLGICLTSTVOLVQLMPYGCILDVHVRNGRL 815  
 DB 175 ATSPRANKELIDEAYVVAAGVSPYVSRLLGICLTSTVOLVQLMPYGCILDVHVRNGRL 234  
 QY 816 GSQDLLNWCQIAGMSYLEVRLVHRDLAARNVLYKSPNNVKITDGLARLDDIDETEV 875  
 DB 235 GSQYLLNWCQIAGMSYLEVRLVHRDLAARNVLYKSPNNVKITDGLARLDDIDETEV 294  
 QY 876 HADGKVPKIMMALESILRRFTHQSDVWSYGVVWELMTFGAPYDGIIPARELPDLLEK 935  
 DB 295 HADGKVPKIMMALESILRRFTHQSDVWSYGVVWELMTFGAPYDGIIPARELPDLLEK 354

QY 936 GERLPDPICITIDVYIMVWKWIDSECRPRELVSEFMAADPDRFVIO-NEDLGP 994  
 DB 355 GERLPDPICITIDVYIMVWKWIDSECRPRELVSEFMAADPDRFVIO-NEDLGP 414  
 QY 995 ASPIDSTFVSLLEDDMDGLVDAEFLVVOGQFCDDPAPAGAGWVHHHRSRSG 1054  
 DB 415 PSPIDSKFRTLMSEEDMEDVDAEFLVPHQGF-----NSPST----- 454  
 QY 1055 GDLTLGLPESEEARPSPL-----APSEGASDVFDGLMGAAKGLQSLPHDPSFLQR 1109  
 DB 455 -----SRTPLLSLSATSNMSTXNCIDRNGG-----H----- 481  
 QY 1110 YSEDPYPLPSETGCVAPLTCSPQPEYVNOCPVPPQPSREGPLPARPAGT-LERA 1168  
 DB 482 -----PVREDFPL-----PAPERVNO-LMPKPKSTAMVQIYVYISLTLATSKL 524  
 QY 1169 KTLSPGKGVVGVDFARFAGAVENPEYL 1195  
 DB 525 PMDSRYON-----SHSTAVDNPEYL 544

## RESULT 15

Q9WVF5

PRELIMINARY;

PRT; 655 AA.

AC Q9WVF5;

DT 01-NOV-1999 (Tremblrel. 12, Created)

DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)

DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)

DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).

GN EGFR.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN (1)

SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=LIVER;

RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C.,

RA Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,

RA Maibach N.J.;

RT "Alternative Transcripts from the Human and Mouse EGFR Gene Encode Carboxy-terminal Truncated Receptors";

RL Submitted (JAN-1999) to the EMBL/Genbank/DBJ databases.

RN (2)

SEQUENCE FROM N.A.

RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVTAJ;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,

RA Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,

RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,

RA Maibach N.J.;

RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Egfr transcripts encoding truncated receptor isoforms";

RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.

RN (3)

SEQUENCE FROM N.A.

RC STRAIN=C57BL/6J; TISSUE=LIVER;

RX MEDLINE=21085660; PubMed=11217851;

RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,

RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,

RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I.,

RA Saito T., Okazaki Y., Gojobori T., Bono H., Kaetawa T., Saito R.,

RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,

RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,

RA Kuehl P., Lewis S., Maeno Y., Nikaide I., Pesole G., Quackenbush J.,

RA Schmitt L.M., Staubli F., Suzuki R., Tomita K., Wagner L., Washio T.,

RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barish G.,

RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,

RA Guernicich S., Hill D., Hofmann M., Hume D.A., Kamita M., Lee N.H.,

RA Lyons P., Marchionni L., Mashima J., Mazzarelli U., Mombereis P.,

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,  
 RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,  
 RA Hayashizaki Y.;

RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690 (2001).  
 DR EMBL: AF124513; ABD44149.1; -  
 DR EMBL: AF275366; AAG28047.1; -  
 DR EMBL: AF275364; AAG28047.1; JOINED.  
 DR EMBL: AF275365; AAG28047.1; JOINED.  
 DR EMBL: AK004944; BAB23688.1; -  
 DR EMBL: AK004883; BAB23641.1; -  
 DR EMBL: AK004911; BAB23662.1; -  
 DR MGI: 95294; Egfr.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR002174; Furin-like.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR SMART: SM00261; FU; 3.  
 KM Receptor.  
 SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

## Query Match

22.1%; Score 1506.5; DB 11; Length 655;

Best Local Similarity 44.5%; Pred. No. 1.9e-105;

Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

QY 11 LILALLPQAA--STQVCTGTDMKLRLPASPTHLDMRLHYOGGVVQGNLELYPTN 68  
 DB 14 LILALLPQAA--STQVCTGTDMKLRLPASPTHLDMRLHYOGGVVQGNLELYPTN 73

QY 69 ASLSFLQDIQEVGVVLAHNOVROVPLRLRIVEGTOLFEDNYALAVLDNGDPLNNTTP 128  
 DB 74 YDLSFLKTIQEVGVVLAHNOVROVPLRLRIVEGTOLFEDNYALAVLDNGDPLNNTTP 124

QY 129 VTGASPGGLRELQSLTETLILKGVYLQGNPOLQVDTLMDI----FKKNGQLATLI 184  
 DB 125 -YGTNRGTGLRELQSLTETLILKGVYLQGNPOLQVDTLMDI----FKKNGQLATLI 180

QY 185 DTVRSRACHPCSPMCKGRSCWSESEDCSLTRTYCAGCA-RCKGPLPTDCHEQCAAG 243  
 DB 181 -QSHPSCKCPDPSCKNGSCWGGGEECKKTKITCAQCSHRCKGRSPDCCHQCAAG 239

QY 244 CTGPKASDLCLAFHNSGICELHCAVLTNTDTFESPNPEGRTFASCTVAPYVY 303  
 DB 240 CTGPKASDLCLAFHNSGICELHCAVLTNTDTFESPNPEGRTFASCTVAPYVY 299

QY 304 LSTDVSCITLVCPLNHCEVTAEADGTORCEKSCPCARVCYGLGWQYIKANSKFGITEL 363  
 DB 300 VVTDHSCYRACQPDYEV-EDGIRKCKCKGCPCKVNGGIGIFK-DTISINATNIK 357

QY 364 -FAGCKKIFGSLAFLESPFDGDPASNTAPLOEQLOVFTLEITGYLYISAMPDLPDL 422  
 DB 358 HFXYCTAIGSDHLIIFVAFKGDSPFTPLPDLRELEITLTYEIGFLLIQAMPDMDTL 417

QY 423 SVFQNLQVIRGRILHNGAVSLTQGGISGLGRSLREKSGALILHNHTHCFVATVP 482  
 DB 418 HAFENLEIRGRKQKQGGFSLAVGLNTSLGRSLKELSDGVYISGRNLCTYANTIM 477

QY 483 DQLFNPHQALLHTANRPDECEVGEGLAQHQLCARGCMGPGPTQCVNCSQFLQOEVE 542  
 DB 478 KKLFGTPNCKTKIMNRARAKDKAVNHVCNPLCSSGCGMPBDVCSQVNSRGRECE 537

QY 543 EGVVLQGLPREVYNANHCPCPEQPOKMGSTCEPEADQCVACAYHDPFCVACRPS 602  
 DB 538 KCVILGEPREFEVNSECICQHECLPQAMNITCTGRGDNCTICAHYIDGHCYCTCA 597

QY 603 GYVPSLSTYPIWKFPEBGACQPCPINTCHSCVDDLDKGC 642  
 DB 598 GINGENNTL-VMKYADANNVCHLCANCTYGCAGPGLQGC 636

Search completed: July 22, 2003, 09:24:29

Job time: 51.75 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 36.875 Seconds

(without alignments)  
4527.811 Million cell updates/sec

Title: SEQ4-653-675-14

Perfect score: 6814  
Sequence: 1 MELALCRWGLLALLPPGA.....TFKGTPTANPEYLGLDVEV 1253

Scoring table: BLOSUM62  
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	* Query Match	Length	DB ID	Description
1	6610	97.0	1255	21	AAV92620 Human heregulin 2
2	6610	97.0	1255	22	AAE12130 Human tyrosine kin
3	6610	97.0	1255	22	AAE60167 HER2 transgene pla
4	6610	97.0	1255	23	AAU74545 Human HER2 (erbB2)
5	6604	96.9	1255	17	AAW01111 HER-2/neu protein.
6	6604	96.9	1255	20	AAV92406 Human HER-2/neu on
7	6604	96.9	1255	21	AAE21198 Human HER-2/neu pr
8	6604	96.9	1255	21	AAE84780 Amino acid sequenc
9	6604	96.9	1255	22	AAE85458 Human HER-2/neu pr
10	6604	96.9	1255	22	AAE88257 HER2/neu amino aci

11	6604	96.9	1255	23	AAE24067 Human Her-2 protei
12	6604	96.9	1255	23	AAE20479 Human Her-2/neu pr
13	6604	96.9	1255	23	AAE51143 Human Her-2/neu on
14	6604	96.9	1255	23	AAU77114 Human Her-2/neu po
15	6604	96.9	1255	23	AAE39568 Sequence of c-erbB
16	6440	94.5	1233	14	AAU96923 Human breast cance
17	6287	92.3	1200	21	AAE21208 Human HER-2/neu pr
18	5821.5	85.4	1256	21	AAE21199 Rat HER-2/neu prot
19	5821.5	85.4	1256	23	AAE51144 Rat Her-2/neu onco
20	5794.5	85.0	1256	23	AAE21206 Mouse Her-2/neu pr
21	5794.5	85.0	1256	22	AAE62860 Amino acid sequenc
22	5794.5	85.0	1256	23	AAE51151 Mouse Her-2/neu on
23	4815	70.7	919	23	AAE21203 Human HER-2/neu fu
24	4815	70.7	919	23	AAE51148 Her-2/neu extracel
25	4065.5	59.7	920	23	AAE51152 Mouse Her-2/neu ex
26	4065.5	59.7	926	23	AAE51153 Mouse Her-2/neu ex
27	3699	54.3	712	23	AAE21204 Human HER-2/neu fu
28	3699	54.3	712	23	AAE51149 Human Her-2/neu fu
29	3552	52.1	782	21	AAE19764 Her2-GW-CSF immuno
30	3550	52.1	653	21	AAE21200 Extracellular HER-
31	3550	52.1	653	23	AAE51145 Human Her-2/neu on
32	3512	51.5	645	22	AAE60408 Human ErbB2 oncogr
33	3512	51.5	645	22	AAE61593 Human ErbB2 extrac
34	3447	50.6	951	21	AAE44993 DC8scfv-erbB2EC fu
35	3344	49.1	624	11	AAE08222 Extracellular port
36	3069	45.0	1210	21	AAE19259 Amino acid sequenc
37	3069	45.0	1210	21	AAE50616 Human EGF receptor
38	3069	45.0	1210	23	AAE23019 Human Her-1 protei
39	3069	45.0	1210	23	AAE50768 Human epidermal gr
40	3067	45.0	1210	22	AAE68420 Amino acid sequenc
41	3053.5	44.8	654	21	AAE21205 Rat HER-2/neu prot
42	3053.5	44.8	654	21	AAE51150 Rat Her-2/neu onco
43	3028	44.4	1210	23	AAE51768 Human epidermal gr
44	2977	43.7	583	23	AAE20483 Human protein for
45	2977	43.7	587	23	AAE20481 Human protein for

## ALIGNMENTS

RESULT 1	AAV92620 standard: Protein; 1255 AA.
ID	AAV92620
AC	AAV92620;
XX	
XX	10-AUG-2000 (first entry)
DT	
XX	
DE	Human heregulin 2 (Her2).
XX	
KW	Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
KW	cell-protein; cancer; breast cancer; prostate cancer;
KW	cell-associated peptide antigen; foreign epitope.
XX	
OS	Homo sapiens.
XX	
XX	
FT	Key
FT	Domain
FT	Location/Qualifiers
FT	1..173
FT	/label= "N-terminal"
FT	/note= "mature polypeptide"
FT	5..25
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	59..73
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	103..117
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	149..163
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	174..323
FT	Domain

FT /label= Cysteine\_rich\_domain  
 FT Region  
 FT 210..224  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 250..264  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 324..483  
 FT /label= Ligand\_binding\_domain  
 FT 325..339  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 369..383  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 465..479  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 484..623  
 FT /label= Cysteine\_rich\_domain  
 FT 579..593  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 624..654  
 FT /label= Transmembrane\_domain  
 FT 632..652  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 653..667  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 655..1010  
 FT /label= Tyrosine\_kinase\_domain  
 FT 661..675  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 695..709  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Region  
 FT 710..730  
 FT /label= insertion\_region  
 FT /note= "suitable for foreign epitope insertion"  
 FT Domain  
 FT 1011..1235  
 FT /label= C-terminal\_domain

WO200020027-A2.  
 13-APR-2000.  
 05-OCT-1999; .99WO-DK00525.  
 05-OCT-1998; 98DK-0001261.  
 20-OCT-1998; 98US-0105011.  
 (MEBI-) M & E BIOTECH AS.  
 Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I,  
 Gautam A, Birk P, Karlsson G;  
 WPI: 2000-349917/30.  
 N-PSDB; AAA09455.  
 Inducing immune responses to weakly immunogenic, tumor associated  
 peptide antigens for the treatment of breast and prostate cancer  
 Claim 62; Page 193-198; 220pp; English.  
 This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of  
 Her2 can be used in the claimed method as an autovaccine to induce a CTL  
 response. Subdominant CTL epitopes, antibody binding regions and  
 cysteine residues involved in disulfide bonds are preserved in the  
 immunogenized forms. Regions suitable for the insertion of foreign T  
 helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic  
 CC cell-associated peptide antigens (PA) such as those associated with  
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen  
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).  
 CC The method comprises effecting simultaneous presentation by antigen  
 CC producing cells (APCs) of the immune system of: (1) at least 1  
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1  
 CC B-cell group derived from the cell-associated PA; and (2) at least 1  
 CC first T helper cell group which is foreign to the animal. Analogues of  
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial  
 CC part of all known and predicted CTL and B-cell epitopes of the respective  
 CC PA and including at least one foreign T helper epitope are also claimed.  
 CC The method is used to treat prostate, prostate/preast or breast cancer  
 CC when the PA is human PSM, FGF8b and Her2, respectively.  
 CC XX

Sequence 1255 AA:

Query Match 97.0%; Score 6610; DB 21; Length 1255;  
 Best Local Similarity 96.9%; Pred. No. 0;  
 Matches 1220; Conservative 10; Mismatches 19; Indels 10; Gaps 2;

QY 1 MEIAAICRMGLLIALPPGAASVQVCTGDMKRLPASPEPTHLMRLHYQGVQVGNL 60  
 DB 1 MEIAAICRMGLLIALPPGAASVQVCTGDMKRLPASPEPTHLMRLHYQGVQVGNL 60  
 QY 61 ELTYLPNASTLFLQDIQEVQGVVLIHANOVROVPLRIYVGTOLFEDNVALAVDNG 120  
 DB 61 ELTYLPNASTLFLQDIQEVQGVVLIHANOVROVPLRIYVGTOLFEDNVALAVDNG 120  
 QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGVYLQRPOLCYQDTILMKDIFHKNNOLA 180  
 DB 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGVYLQRPOLCYQDTILMKDIFHKNNOLA 180  
 QY 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGVYLQRPOLCYQDTILMKDIFHKNNOLA 180  
 DB 121 DPLNNTTPTVGTASPGGLRELQRLSLTEILKGVYLQRPOLCYQDTILMKDIFHKNNOLA 180  
 QY 181 LTLIDPTRSRAPCHPCSMCKSGRCWSSSDCSLRTVAGGACARCKGLPDDCCEQC 240  
 DB 181 LTLIDPTRSRAPCHPCSMCKSGRCWSSSDCSLRTVAGGACARCKGLPDDCCEQC 240  
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMNPPEGRTYFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMNPPEGRTYFGASCVTACP 300  
 QY 301 YNYISTVSGCTLVCPAHNOEYTAEDGTQCEKSCRCACVCGLMQVYIKANSKFTGIT 360  
 DB 301 YNYISTVSGCTLVCPAHNOEYTAEDGTQCEKSCRCACVCGLMQVYIKANSKFTGIT 360  
 QY 361 ELBPAQCKITFGSLAFPESEFDDPASNTPALQPEQLVETLEITGYLYISAMPDPLP 420  
 DB 361 IQBPAQCKITFGSLAFPESEFDDPASNTPALQPEQLVETLEITGYLYISAMPDPLP 420  
 QY 421 DLSVFQNLQVIRGILHNGAYSLTLOGLISWIGLSRLSGGLALIHNTHLCPVHTV 480  
 DB 421 DLSVFQNLQVIRGILHNGAYSLTLOGLISWIGLSRLSGGLALIHNTHLCPVHTV 480  
 QY 481 PMDQLEPNPQALHTANRPEDCEVGGGLACHOLCAAGHMGSGPTQVNCOSPLRQEC 540  
 DB 481 PMDQLEPNPQALHTANRPEDCEVGGGLACHOLCAAGHMGSGPTQVNCOSPLRQEC 540  
 QY 541 VEECRVLOGLPREVYNAHCLPCHPECQPNQSVTCGPEADQCVACAHYKDPFCVARC 600  
 DB 541 VEECRVLOGLPREVYNAHCLPCHPECQPNQSVTCGPEADQCVACAHYKDPFCVARC 600  
 QY 601 PSQVPRDLSYMPIMKPFDEEGACQPCINCTHSCVDLDDGCPAEOGASPLTISVAVVG 660  
 DB 601 PSQVPRDLSYMPIMKPFDEEGACQPCINCTHSCVDLDDGCPAEOGASPLTISVAVVG 660  
 QY 661 ILVVVLGVVFGILI-----FNNFTVSWLRVVKYSASHLEPLTBSGAMPNOAKMILK 714  
 DB 661 ILVVVLGVVFGILI-----FNNFTVSWLRVVKYSASHLEPLTBSGAMPNOAKMILK 714  
 QY 715 ETELRKRVKUGSAGFGVYVGIWIPDSGNKIPVATVLAENSPKANKELIDBAYVMAG 774  
 DB 715 ETELRKRVKUGSAGFGVYVGIWIPDSGNKIPVATVLAENSPKANKELIDBAYVMAG 774

QY 775 VGSPPVSRLLGICLTSTVQLVQLMPYGLLDHVRNRRGLSGSOLLNMCQAKGMSYL 834  
 PS 777 VGSPPVSRLLGICLTSTVQLVQLMPYGLLDHVRNRRGLSGSOLLNMCQAKGMSYL 836  
 Db 835 EDVRLVHRDLAARNVLKSPNHYKIDTFGLARLLIDETVHADGKVPKXKALSLIR 894  
 Db 837 EDVRLVHRDLAARNVLKSPNHYKIDTFGLARLLIDETVHADGKVPKXKALSLIR 896  
 QY 895 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIDPLEKGERLPPPICTIDVYMINV 954  
 Db 897 RRFTHQSDVWSYGVTVWELMTFGAKPYDGIIPAREIDPLEKGERLPPPICTIDVYMINV 956  
 QY 955 KCMWIDSECRPRRELVSERSRMARDPQRFVTVQNEBGLASPJLSTFYRSLIEDDDMGD 1014  
 Db 957 KCMWIDSECRPRRELVSERSRMARDPQRFVTVQNEBGLASPJLSTFYRSLIEDDDMGD 1016  
 QY 1015 LVDAEEYLVPOQGFCCDPAPAGAGMWHHRSSSTRSGGDLTLGLPEEERAPSPILA 1074  
 Db 1017 LVDAEEYLVPOQGFCCDPAPAGAGMWHHRSSSTRSGGDLTLGLPEEERAPSPILA 1076  
 QY 1075 PSEGAGSDVDFDGLGMAAGLQSLPTHDSPLQRYSEDETVLPSSETDGYVAPLTCSPQ 1134  
 Db 1077 PSEGAGSDVDFDGLGMAAGLQSLPTHDSPLQRYSEDETVLPSSETDGYVAPLTCSPQ 1136  
 QY 1135 PEYVNOQDVPRQPPSPPEEGPLPAARPGATLERAKTSLPGKNGVWVDVAFGAVENPEY 1194  
 Db 1137 PEYVNOQDVPRQPPSPPEEGPLPAARPGATLERAKTSLPGKNGVWVDVAFGAVENPEY 1196  
 QY 1195 LTFQGAAPQPHPPAPSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYGLDVPV 1253  
 Db 1197 LTFQGAAPQPHPPAPSPAFDNLVYWDQDPPERGAAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 2  
 AAE12130  
 ID AAE12130 standard; Protein: 1255 AA.  
 AC AAE12130;  
 DT 18-DEC-2001 (first entry)  
 DE Human tyrosine kinase-type receptor, HER-2.  
 KW Therapeutic compound; major histocompatibility complex; vaccine;  
 KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;  
 KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;  
 KW antigen presenting cell; human; tyrosine kinase-type receptor.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Region 774..782  
 FT /note="Antigenic epitope"  
 XX WO200168677-A2.  
 XX 20-SEP-2001.  
 XX 16-MAR-2001; 2001MO-US40328.  
 XX 16-MAR-2000; 2000US-0527487.  
 XX (GENZ ) GENZYME CORP.  
 XX Nicolette CA;  
 XX WBI: 2001-616284/71.  
 XX N-PSDB; AAD19731.  
 PT Novel synthetic therapeutic compound for inducing immune response and  
 PT for use in adoptive immunotherapy, has enhanced binding to major  
 PT histocompatibility molecules and enhanced immunoregulatory properties

XX Claim 4; Page 63-67; 69pp; English.  
 PS The invention relates to synthetic therapeutic compounds (antigenic  
 CC peptides) with enhanced binding to major histocompatibility complex  
 CC (MHC) molecules and enhanced immunoregulatory properties relative  
 CC to their natural counterparts. Compounds of the invention are useful  
 CC for inducing an immune response in a subject and for use in adoptive  
 CC immunotherapy. They are useful as components of anti-cancer vaccines  
 CC and to expand immune effector cells that are specific for cancers  
 CC characterized by expression of the breast cancer antigen, HER-2.  
 CC Polynucleotides that encode peptides of the invention are useful as  
 CC hybridization probes and as primers for the detection of genes of gene  
 CC transcripts that are expressed in antigen presenting cells (APCs), to  
 CC confirm transduction of polynucleotides into host cells. The present  
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds  
 CC of the invention are designed based on the HER-2 antigenic peptide  
 CC (774-782).  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 97.0%; Score 6610; DB 22; Length 1255;  
 Best Local Similarity 96.9%; Pred. No. 0;  
 Matches 1220; Conservative 10; Mismatches 19; Indels 10; Gaps 2;  
 QY 1 MELAALCRWGLLIALLPQAASTQVCTGDMKRLPASPEHLDLRLHLYGCGVYQGNL 60  
 Db 1 MELAALCRWGLLIALLPQAASTQVCTGDMKRLPASPEHLDLRLHLYGCGVYQGNL 60  
 QY 61 ELTYLPTNALSFLDDIOEVQGVYLAHNOYRQVLOGLRIYRGQLFEDNYALAVLNG 120  
 Db 61 ELTYLPTNALSFLDDIOEVQGVYLAHNOYRQVLOGLRIYRGQLFEDNYALAVLNG 120  
 QY 121 DELNNTPTVGTASPGGLRELOLRSLTEILKGVLIQRNPOLCYQDTILMKDIFHNQOLA 180  
 Db 121 DELNNTPTVGTASPGGLRELOLRSLTEILKGVLIQRNPOLCYQDTILMKDIFHNQOLA 180  
 QY 181 LTLIDTNRBRACHPSCPMCKSRCKWSESSDQCSTLRVYCAAGCARGPLPTDCCHQC 240  
 Db 181 LTLIDTNRBRACHPSCPMCKSRCKWSESSDQCSTLRVYCAAGCARGPLPTDCCHQC 240  
 QY 241 AAGCTGPRKSDCLALCFHNSGICELHCPALVTVYTDFFESMPNDEGRYTGASCVTACP 300  
 Db 241 AAGCTGPRKSDCLALCFHNSGICELHCPALVTVYTDFFESMPNDEGRYTGASCVTACP 300  
 QY 301 NVYISTDVSGCTVCPHNOEYTAEDGTQRCCKSKPCARVCYGLGMYIKANSKFIGIT 360  
 Db 301 NVYISTDVSGCTVCPHNOEYTAEDGTQRCCKSKPCARVCYGLGMYIKANSKFIGIT 360  
 QY 361 ELIEFAGCKKIRGSLAFLPESPFGDPSNTAPLOPQLQVFETLEITGVLYISAMPDLSL 420  
 Db 361 IOEFAGCKKIRGSLAFLPESPFGDPSNTAPLOPQLQVFETLEITGVLYISAMPDLSL 420  
 QY 421 DLSVFQNLQVIRGRILHNGAVSLTQGLGIMLGLRSLRELSGALILHNHTLFCFATV 480  
 Db 421 DLSVFQNLQVIRGRILHNGAVSLTQGLGIMLGLRSLRELSGALILHNHTLFCFATV 480  
 QY 481 PWDQLFRRPHQALLHTARPEDECVGBGLAQHQLCARHCGPPTQCVNCSQFLRGQEC 540  
 Db 481 PWDQLFRRPHQALLHTARPEDECVGBGLAQHQLCARHCGPPTQCVNCSQFLRGQEC 540  
 QY 541 VEECRVLQGLPREYVNAHCLPCHBECOPONGSVTCFPEADQCVACHYADPPCVARC 600  
 Db 541 VEECRVLQGLPREYVNAHCLPCHBECOPONGSVTCFPEADQCVACHYADPPCVARC 600  
 QY 601 PSGVPRDLSYMPIMKFPDEBGAQCPPINCHSCVLDLDCQCPAQRASPLTISYSAVG 660  
 Db 601 PSGVPRDLSYMPIMKFPDEBGAQCPPINCHSCVLDLDCQCPAQRASPLTISYSAVG 660  
 QY 661 ILVVVLGVVFGIIL-----FNNTVAFWLRVPVSASHLEPLTPSGAMPQOQMRILK 714  
 Db 661 ILVVVLGVVFGIILKRNQKIRKTYMKRLLOETEL---VEPLTPSGAMPQOQMRILK 716



```

QY 715 ETELARKVYLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVMAG 774
Db 717 ETELARKVYLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVMAG 776
QY 775 VGSPIVSRLLIGICLTSTVQLVTOQMPYGCGLDHRNRRGRGLSGSODLNMCMQIAKGMSTL 834
Db 777 VGSPIVSRLLIGICLTSTVQLVTOQMPYGCGLDHRNRRGRGLSGSODLNMCMQIAKGMSTL 836
QY 835 EDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGKVPFKMALESILR 894
Db 837 EDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGKVPFKMALESILR 896
QY 895 RRFTHQSDVMSYGVTVMEIMTFGAKPXDGIIPAREIPDLTXGERLPQPICTIDVYIMV 954
Db 897 RRFTHQSDVMSYGVTVMEIMTFGAKPXDGIIPAREIPDLTXGERLPQPICTIDVYIMV 956
QY 955 KCMWIDSECRPRFRELVSFSSRMARDPQRFVVIQNEDIGPASPILDSTFYRSLLEDMDMGD 1014
Db 957 KCMWIDSECRPRFRELVSFSSRMARDPQRFVVIQNEDIGPASPILDSTFYRSLLEDMDMGD 1016
QY 1015 LVDAEYLVPOQGFPCDPAPAGAMVHRRSSSTRSGGDLTLGLEPSEEAPRSPLA 1074
Db 1017 LVDAEYLVPOQGFPCDPAPAGAMVHRRSSSTRSGGDLTLGLEPSEEAPRSPLA 1076
QY 1075 PSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOYSBDPTVPLPSETDGYVAPLTCSPQ 1134
Db 1077 PSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOYSBDPTVPLPSETDGYVAPLTCSPQ 1136
QY 1135 PEYVNPQPVRRPQPSPREGRLPARAPAGATLEAKTJLSPGNVGVKVFAPAGAVENPEY 1194
Db 1137 PEYVNPQPVRRPQPSPREGRLPARAPAGATLEAKTJLSPGNVGVKVFAPAGAVENPEY 1196
QY 1195 LTPQGAAPQPPPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1253
Db 1197 LTPQGAAPQPPPPAFSPAFDNLVYWDQDPPERGAPSTFKGTPTAENPEYGLDVPV 1255

```

RESULT 3  
AAB60167  
ID AAB60167 standard; Protein; 1255 AA.  
AC AAB60167;  
DT 03-APR-2001 (first entry)  
XX HER2 transgene plasmid construct encoded protein.  
XX Human; HER2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;  
XX antibody.  
XX Homo sapiens.  
XX Synthetic.  
XX WO200100244-A2.  
XX 04-JAN-2001.  
XX 23-JUN-2000; 2000WO-US172229.  
XX 25-JUN-1999; 99US-0141316.  
XX 16-MAR-2000; 2000US-0189844.  
XX (GETH ) GENENTECH INC.  
XX Erickson S, Schwall R;  
XX WPI: 2001-061962/07.  
XX N-P3DB; AAF24297.  
XX Treating tumors, particularly breast cancers, which overexpress an ErbB  
XX receptor and does not respond to an anti-ErbB antibody, comprises  
XX conjugating the antibody to a maytansinoid -

```

PS Example 3; Fig 4; 92pp; English.
XX
CC The present invention provides a method of treating cancer by
CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In
CC particular, the antibody is directed against ErbB2 (also known as HER2
CC and p185neu). The method is particularly useful in the treatment of
CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,
CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.
XX
SQ Sequence 1255 AA;
Query Match 97.0%; Score 6610; DB 22; Length 1255;
Best Local Similarity 96.9%; Pred. No. 0;
Matches 1220; Conservative 10; Mismatches 19; Indels 10; Gaps 2;
QY 1 MELIALCRWGLLALLPFGAASVQVCTGDMKLRPASBETHLDMRLHYQCGVQVQNL 60
Db 1 MELIALCRWGLLALLPFGAASVQVCTGDMKLRPASBETHLDMRLHYQCGVQVQNL 60
QY 61 ELTYLPFNASLSFLDIOEVQGVYLIHNOVROVPLQRIYRGTLQFEDNYALAVLDNG 120
Db 61 ELTYLPFNASLSFLDIOEVQGVYLIHNOVROVPLQRIYRGTLQFEDNYALAVLDNG 120
QY 121 DPLNNTPTVYGASPGGLRELDLSLTELKGGVLIQRNPQLCYODTILMKDIFHXNQOLA 180
Db 121 DPLNNTPTVYGASPGGLRELDLSLTELKGGVLIQRNPQLCYODTILMKDIFHXNQOLA 180
QY 181 LTLIDPNRRACGPGSPMKCKGSRGWSESEDDQSILRTYCAGGACAKCPPLPDDCHEOC 240
Db 181 LTLIDPNRRACGPGSPMKCKGSRGWSESEDDQSILRTYCAGGACAKCPPLPDDCHEOC 240
QY 241 AAGCTGPKXSDCLACIHFHNSGICELHCALVTYNTDRESMPNPEGRTPFGASCYACP 300
Db 241 AAGCTGPKXSDCLACIHFHNSGICELHCALVTYNTDRESMPNPEGRTPFGASCYACP 300
QY 301 YNLTSTDVSGCTLVCLHNOEVTAEEDGTORCEKSPCARVYCYGLMOYIKANSKIGIT 360
Db 301 YNLTSTDVSGCTLVCLHNOEVTAEEDGTORCEKSPCARVYCYGLMOYIKANSKIGIT 360
QY 361 ELIEFACCKIFGSLAFPSFDGDPASNTAPQPEOLQVFEETLEETGLYISAMPDSL 420
Db 361 IOEPACCKIFGSLAFPSFDGDPASNTAPQPEOLQVFEETLEETGLYISAMPDSL 420
QY 421 DLSVFQNLQVIRGRIIHNGAYSILTLQGLISWGLRSLELSGGLLIHNNHLCVHTV 480
Db 421 DLSVFQNLQVIRGRIIHNGAYSILTLQGLISWGLRSLELSGGLLIHNNHLCVHTV 480
QY 481 PMDQLFRNPHQALHTANRPEDECVSEGLAQHLCARGHGWGPGPQCVNCSOFLRGQEC 540
Db 481 PMDQLFRNPHQALHTANRPEDECVSEGLAQHLCARGHGWGPGPQCVNCSOFLRGQEC 540
QY 541 VEECRVLOGLPREYVNAHCLPCHPEQOPQNSVITCFGEADQCVACAHYKDPFCVABC 600
Db 541 VEECRVLOGLPREYVNAHCLPCHPEQOPQNSVITCFGEADQCVACAHYKDPFCVABC 600
QY 601 PSGVRPDLSTYMPITWKFPDEEGACQPCPINCTHSVDLDKGCAPAEPRAPLTSYSAVVG 660
Db 601 PSGVRPDLSTYMPITWKFPDEEGACQPCPINCTHSVDLDKGCAPAEPRAPLTSYSAVVG 660
QY 661 ILLVVLGVVFGILL-----FNNFTVSFMLEVPKVSASHLEPLTPSGAMPNQAQMRILK 714
Db 661 ILLVVLGVVFGILLIRROOKIRKYMRLDLQETL-----VEPLTPSGAMPNQAQMRILK 716
QY 715 ETELARKVYLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVMAG 774
Db 717 ETELARKVYLGSGAFGTYYKGIWIPDGENVKIPVAIKVLRNTSPKANKELIDEAYVMAG 776
QY 775 VGSPIVSRLLIGICLTSTVQLVTOQMPYGCGLDHRNRRGRGLSGSODLNMCMQIAKGMSTL 834
Db 777 VGSPIVSRLLIGICLTSTVQLVTOQMPYGCGLDHRNRRGRGLSGSODLNMCMQIAKGMSTL 836
QY 835 EDVRLVHRDLAARNVLVKSPPNHVKITDFGLARLLDIDETEHADGKVPFKMALESILR 894

```

Db 837 EDVRLVARDLAARNVLVKSPNHVKITDPLGLALDLIDETEHADGKVPKXMALESILR 896  
 QY 895 RRETHOSDWSYGVTVWELMTFGAKPYDGI.PAREIPDLLEKGRLEPPICITIDYVIMW 954  
 Db 897 RRTTHOSDWSYGVTVWELMTFGAKPYDGI.PAREIPDLLEKGRLEPPICITIDYVIMW 956  
 QY 955 KCMWIDSECRPRFRELSEFSRMDPQRFVVI.QNEDLGPASPLDSTFYSLLEDDMDG 1014  
 Db 957 KCMWIDSECRPRFRELSEFSRMDPQRFVVI.QNEDLGPASPLDSTFYSLLEDDMDG 1016  
 QY 1015 LVDAEYLVPQQGFFCCDDPAPAGAGWVHHHRSSSTSSGGGDLTLGLEBEEBAPSPLA 1074  
 Db 1017 LVDAEYLVPQQGFFCCDDPAPAGAGWVHHHRSSSTSSGGGDLTLGLEBEEBAPSPLA 1076  
 QY 1075 PSEGAGSDVFDGDLGMAAAGLQSLPTHDSPLQRYSESDTVPLPSETDGVVPLTCSFQ 1134  
 Db 1077 PSEGAGSDVFDGDLGMAAAGLQSLPTHDSPLQRYSESDTVPLPSETDGVVPLTCSFQ 1136  
 QY 1135 PEYVNOPDVPRQPPSPREGPLPAARPAATLERAKTILSPKNGVYKDVFAFGAVENPEY 1194  
 Db 1137 PEYVNOPDVPRQPPSPREGPLPAARPAATLERAKTILSPKNGVYKDVFAFGAVENPEY 1196  
 QY 1195 LTPQGGAPOPHPPAPSPAFDNLXYWDCOPPERGAPSPSTFKGPTTENPEYLGIVPV 1253  
 Db 1197 LTPQGGAPOPHPPAPSPAFDNLXYWDCOPPERGAPSPSTFKGPTTENPEYLGIVPV 1255  
 RESULT 4  
 AAU74545  
 ID AAU74545 standard; Protein; 1255 AA.  
 AC AAU74545;  
 XX  
 XX 23-Apr-2002 (first entry)  
 DE Human HER2 (ErbB2) polypeptide.  
 XX  
 XX Human, HER2; ErbB; epidermal growth factor receptor; receptor;  
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;  
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;  
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;  
 KW glioma disorder; astrocytoma disorder; hypothalamic disorder;  
 KW glandular disorder; macrophagal disorder; epithelial disorder;  
 KW stromal disorder; blastocellic disorder; inflammatory disorder;  
 KW angiogenic disorder; immunological disorder.  
 XX  
 XX Homo sapiens.  
 OS  
 PN US2002001587-A1.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 16-MAR-2001; 2001US-0811123.  
 XX  
 PR 16-MAR-2000; 2000US-189844P.  
 PR 05-OCT-2000; 2000US-238327P.  
 XX  
 PA (ERIC/) ERICKSON S.  
 PA (SCHW/) SCHWALL R.  
 PA (SLIM/) SLIKOWSKI M.  
 XX  
 PI Erickson S, Schwall R, Slikowski M;  
 XX  
 DR WPI; 2002-163686/21.  
 DR N-PSDB; ABK14058.  
 XX  
 PT Treating tumour characterised by overexpression of epidermal growth  
 PT factor receptor, ErbB or cancer in mammal, comprises administering  
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal  
 XX  
 PS Example 3; Fig 7; 93pp; English.  
 XX  
 CC The invention relates to treating a tumour in a mammal, where the tumour

CC is characterised by the overexpression of an epidermal growth factor  
 CC receptor (ErbB) and does not respond or responds poorly, to treatment  
 CC with an anti-ErbB antibody, comprising administering to the mammal an  
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for  
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,  
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,  
 CC prostate and bladder, preferably breast cancer. The breast cancer is a  
 CC metastatic breast cancer or an aggressive form of metastatic breast  
 CC cancer which overexpresses ErbB2. The method is also useful for treating  
 CC neuronal, glial, astrocytes, hypothalamic, glandular, macrophagal,  
 CC epithelial, stromal, blastocellic, inflammatory, angiogenic and  
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)  
 CC polypeptide of the invention.  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 97.0%; Score 6610; DB 23; Length 1255;  
 Best Local Similarity 96.9%; Pred. No. 0;  
 Matches 1220; Conservative 10; Mismatches 19; Indels 10; Gaps 2;  
 QY 1 METALCRGILLALPPGAATGYCTGDMTLRLPASETHLDMRLHYGCGQVYQGNL 60  
 Db 1 MELALCRGILLALPPGAATGYCTGDMTLRLPASETHLDMRLHYGCGQVYQGNL 60  
 QY 61 ELTYLPTNASLFLDIOEVGGVYLIANHQRVPLQRLIRIVRGTLQFEDYALAVLNG 120  
 Db 61 ELTYLPTNASLFLDIOEVGGVYLIANHQRVPLQRLIRIVRGTLQFEDYALAVLNG 120  
 QY 121 DPLNNTPTVTGASPGGLRELQRLSTELLKGVGLIQRNQLCYQDTILMKDIFHKNNOLA 180  
 Db 121 DPLNNTPTVTGASPGGLRELQRLSTELLKGVGLIQRNQLCYQDTILMKDIFHKNNOLA 180  
 QY 181 LTLIDTNRSPACHPSPCMKSGRCWSESEDCSLRTVCAGACARCKPLPTDCHEQC 240  
 Db 181 LTLIDTNRSPACHPSPCMKSGRCWSESEDCSLRTVCAGACARCKPLPTDCHEQC 240  
 QY 241 AAGCTGPKMSDCLACHFNHSGICELHCAVLTNTDTFESPNPNEGRTFASCTYAC 300  
 Db 241 AAGCTGPKMSDCLACHFNHSGICELHCAVLTNTDTFESPNPNEGRTFASCTYAC 300  
 QY 301 VVYLSIDVSGCTVCPLEHNOEYTAEDGTORCKSKPCARVCYGGMOYIKANSKFIGIT 360  
 Db 301 VVYLSIDVSGCTVCPLEHNOEYTAEDGTORCKSKPCARVCYGGMOYIKANSKFIGIT 360  
 QY 361 ELBPAQCKKIFGSLAFLPESFDGPASNTAPLQPSQLQVFETLEBITGYLIYSAMPDILP 420  
 Db 361 IOBPAQCKKIFGSLAFLPESFDGPASNTAPLQPSQLQVFETLEBITGYLIYSAMPDILP 420  
 QY 421 DLSVQNTQVIRGLIHNGAVSLTQGLISWLGSRSLREISGGLALIHNNHLCFVHTV 480  
 Db 421 DLSVQNTQVIRGLIHNGAVSLTQGLISWLGSRSLREISGGLALIHNNHLCFVHTV 480  
 QY 481 PMDOLFRRNHQALLHTANRPEDECVGEGIALCHQLCARHGWPGFTQCVNCSQFLRGQEC 540  
 Db 481 PMDOLFRRNHQALLHTANRPEDECVGEGIALCHQLCARHGWPGFTQCVNCSQFLRGQEC 540  
 QY 541 VEECRVLOGLPREYVNAHCLPCHCEQOPONSVTGCFGEPAQCVAACHYKXPPCVNAC 600  
 Db 541 VEECRVLOGLPREYVNAHCLPCHCEQOPONSVTGCFGEPAQCVAACHYKXPPCVNAC 600  
 QY 601 PSGVPRDLSYMPYIKFPEDEGACQCPINCHSCVDLDDKGPAPQBSPLTSIYSAVVG 660  
 Db 601 PSGVPRDLSYMPYIKFPEDEGACQCPINCHSCVDLDDKGPAPQBSPLTSIYSAVVG 660  
 QY 661 ILVVVVLGVVFGILT-----FNNFTVSFWLRVPPVYASASHLEPLTPSGAMPQAOAMRLK 714  
 Db 661 ILVVVVLGVVFGILT-----FNNFTVSFWLRVPPVYASASHLEPLTPSGAMPQAOAMRLK 714  
 QY 715 ETELKRVKVLGAGAGTYKGIWPDGENVKIPVAKYLRNTSKAKKEIDDEAYVWAG 774  
 Db 715 ETELKRVKVLGAGAGTYKGIWPDGENVKIPVAKYLRNTSKAKKEIDDEAYVWAG 774  
 QY 775 VGSPPVSRLLGICLTSTVQVLTQMLPYGCLLDHVENRNGRLGSGODLNMWQIAKGMGYL 834  
 Db 775 VGSPPVSRLLGICLTSTVQVLTQMLPYGCLLDHVENRNGRLGSGODLNMWQIAKGMGYL 834

```

DB 777 VGSFVSRLLIGICTSTVQLVTLQMLPGCLLDHVERGRGLSGDILNMCQIAKMSYL 836
QY 835 EDVFLVHRDLAARVAVKSPVHVKITPFGGLARLDIDETETHADGGKVPFKMMALBSILR 894
DB 837 EDVFLVHRDLAARVAVKSPVHVKITPFGGLARLDIDETETHADGGKVPFKMMALBSILR 896
QY 895 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKEEFLPPPICTIDVYIMV 954
DB 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKEEFLPPPICTIDVYIMV 956
QY 955 KCMWIDSECRPFREIUSEFSRMAADPREFVIONEDLGPASPILDSFFYSILJEDDMD 1014
DB 957 KCMWIDSECRPFREIUSEFSRMAADPREFVIONEDLGPASPILDSFFYSILJEDDMD 1016
QY 1015 LVDAEYLVPOQGFECDDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSPLA 1074
DB 1017 LVDAEYLVPOQGFECDDPAPAGAGMWHRRSSSTRSGGDLTLGLEPSEEAAPRSPLA 1076
QY 1075 PSEGAGSTVFPGDILGMAKAGLSLPTHDSPLQRYSEDPVPLPSETDGVVAPLTCSFQ 1134
DB 1077 PSEGAGSTVFPGDILGMAKAGLSLPTHDSPLQRYSEDPVPLPSETDGVVAPLTCSFQ 1136
QY 1135 PEYVNPQPVRRPQPSPREGGLPAPAPAGATLERAKTLSPGKGVVXDVFAFGAVENPEY 1194
DB 1137 PEYVNPQPVRRPQPSPREGGLPAPAPAGATLERAKTLSPGKGVVXDVFAFGAVENPEY 1196
QY 1195 LTPQGGAAPQHPHPPAPFADNLYYNDQDPPEGAPSTFKGPTLHNPBYLGLDVPV 1253
DB 1197 LTPQGGAAPQHPHPPAPFADNLYYNDQDPPEGAPSTFKGPTLHNPBYLGLDVPV 1255

```

RESULT 5  
AAM01111 standard; Protein, 1255 AA.

```

ID AAM01111
AC AAM01111;
DT 01-JAN-1997 (first entry)
DE HER-2/neu protein.
XX HER-2/neu protein.
XX HER-2/neu; c-erbB1, p185, oncogene; tyrosine protein kinase;
XX breast cancer; ovary cancer; colon cancer; lung cancer;
XX prostate cancer; immunisation; tumour; vaccine; vector.
XX OS Homo sapiens.
XX FH Key Location/Qualifiers
XX FT Domain 676..1255
XX FT /label= intracellular_domain
XX FT /note= "claimed domain, useful for immunisation"
XX PN MO9630514-A1.
XX PD 03-OCT-1996.
XX PF 28-MAR-1996; 96WO-US01689.
XX PR 31-MAR-1995; 95US-0414417.
XX PA (UNITM) UNIV WASHINGTON.
XX PI Cheever MA, Disis ML;
XX WPI; 1996-455361/45.
XX DR N-PSDB; AAT40739.
XX DNA encoding HER-2-neu poly:peptide(s) - used for prevention or
XX treatment of malignancies with which the HER-2/neu oncogene is
XX associated
PS Claim 2; Page 56-61; 71pp; English.

```

```

XX Human HER-2/neu protein (AAM01111), also called p185 or c-erbB2, is
CC the product of the HER-2/neu oncogene (see also AAT40739). The
CC protein is over-expressed in various cancers, including breast,
CC ovarian, colon, lung and prostate. The intracellular domain of the
CC protein can be used to immunise an animal against a malignancy with
CC which the oncogene is associated. The polypeptide can be produced
CC in transformed host cells for use in immunisation. Alternatively,
CC animal cells are transfected in vivo or ex vivo with a viral vector
CC that directs expression of the polypeptide.
XX
SQ Sequence 1255 AA;

```

```

Query Match 96.9%; Score 6604; DB 17; Length 1255;
Best Local Similarity 96.7%; Pred. No. 0;
Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2.

QY 1 METALCRGGLLALLPAASTQVCTGDMTLRLPASBETHLDMRLHYQCQVQGNL 60
DB 1 METALCRGGLLALLPAASTQVCTGDMTLRLPASBETHLDMRLHYQCQVQGNL 60
QY 61 ELTYLPTNASLSFLQDIOEVQGVLIANQVRQVPLQRLIRVGTQLFEDNYALVLDNG 120
DB 61 ELTYLPTNASLSFLQDIOEVQGVLIANQVRQVPLQRLIRVGTQLFEDNYALVLDNG 120
QY 121 DPLNNTPTVPGASPGGLRELQRLSTELIKGVLLQGNPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTPTVPGASPGGLRELQRLSTELIKGVLLQGNPOLCYODTILMKDIFHKNNOLA 180
QY 181 LTLIDNRSRACHPCSPMKCGSRCWGESSEDCSLTRVYACAGCARCKPLPTDCHEQC 240
DB 181 LTLIDNRSRACHPCSPMKCGSRCWGESSEDCSLTRVYACAGCARCKPLPTDCHEQC 240
QY 241 AAGCTPKHSDCLACHFNHSGICEHCPALVTYNTDFEESMPNPEGRTPFASCTAACP 300
DB 241 AAGCTPKHSDCLACHFNHSGICEHCPALVTYNTDFEESMPNPEGRTPFASCTAACP 300
QY 301 YNLTSTDVSGCTLVCELIHQEYTAEDGTQRCCKSPCARVCYGLGMQYIKANSKEITGT 360
DB 301 YNLTSTDVSGCTLVCELIHQEYTAEDGTQRCCKSPCARVCYGLGMQYIKANSKEITGT 360
QY 361 ELFFACCKXIFGSLALPESFGDDPASNAPLOPELOVFEETLEETGLYISAMPDISLP 420
DB 361 ELFFACCKXIFGSLALPESFGDDPASNAPLOPELOVFEETLEETGLYISAMPDISLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSITLQGLISWLGRLSRELSSGALLIHNTHLCVHTV 480
DB 421 DLSVFQNLQVIRGRILHNGAYSITLQGLISWLGRLSRELSSGALLIHNTHLCVHTV 480
QY 481 PMDQFRNPHQALLHTANPEDECVGEGJACHQLCARGHCMGPGPTQCVNCSQFLRGQEC 540
DB 481 PMDQFRNPHQALLHTANPEDECVGEGJACHQLCARGHCMGPGPTQCVNCSQFLRGQEC 540
QY 541 VEECRVLQGLPREYVVARHCLPCHPECOFONGSVTCFGEADQCVACAHYKPPFCVAC 600
DB 541 VEECRVLQGLPREYVVARHCLPCHPECOFONGSVTCFGEADQCVACAHYKPPFCVAC 600
QY 601 PSGVKPDLSTYMPIWKPEDEGACQPCPINCTHSCVDLDDKGPAGGRASPITSISAVVG 660
DB 601 PSGVKPDLSTYMPIWKPEDEGACQPCPINCTHSCVDLDDKGPAGGRASPITSISAVVG 660
QY 661 ILLVVLGVVFGILLI-----FNNFVSFWLAVPKYSASHLEPLTSSGAMPQAOQRILK 714
DB 661 ILLVVLGVVFGILLI-----FNNFVSFWLAVPKYSASHLEPLTSSGAMPQAOQRILK 714
QY 715 ETELARKVYLGSAGFTVYKGIWIPDGENVKIIPVALKVLRENTSPKANKELIDEAVVAG 774
DB 715 ETELARKVYLGSAGFTVYKGIWIPDGENVKIIPVALKVLRENTSPKANKELIDEAVVAG 774
QY 775 VGSFVSRLLIGICTSTVQLVTLQMLPGCLLDHVERGRGLSGDILNMCQIAKMSYL 836
DB 775 VGSFVSRLLIGICTSTVQLVTLQMLPGCLLDHVERGRGLSGDILNMCQIAKMSYL 836

```

QY 835 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGKVPIMKMALESILR 894  
 DB 837 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGKVPIMKMALESILR 896  
 QY 895 RRFTHQSDVMSYGTWELMTFGAKPYDGIIPAREIPDLLENGERLPPQPICTIDVYIMV 954  
 DB 897 RRFTHQSDVMSYGTWELMTFGAKPYDGIIPAREIPDLLENGERLPPQPICTIDVYIMV 956  
 QY 955 KCMWIDSECRPRFELVSEFSRMARDPQRFVIVIONEDLGPAISPLDSTFYRSLLEDMDGD 1014  
 DB 957 KCMWIDSECRPRFELVSEFSRMARDPQRFVIVIONEDLGPAISPLDSTFYRSLLEDMDGD 1016  
 QY 1015 LVDAEEYLVPOQGFCCPPAPAGAGMHHRRSSSTSGGDDLTLGLPSEBEAPRSPLA 1074  
 DB 1017 LVDAEEYLVPOQGFCCPPAPAGAGMHHRRSSSTSGGDDLTLGLPSEBEAPRSPLA 1076  
 QY 1075 PSEGASDVFPGDGLGMAKAGLSLPTHPSPLORYSEDPVPLPSETDGVAPLTCSPQ 1134  
 DB 1077 PSEGASDVFPGDGLGMAKAGLSLPTHPSPLORYSEDPVPLPSETDGVAPLTCSPQ 1136  
 QY 1135 PEYVNPDPVRPQPSPREGPPAPAPAGATLERAKTSLPGKNGVYKDVAFAGAVENPEY 1194  
 DB 1137 PEYVNPDPVRPQPSPREGPPAPAPAGATLERAKTSLPGKNGVYKDVAFAGAVENPEY 1196  
 QY 1195 LTPGGAGAPQHPHPPAPASPAFDNLVYWDQDPERGAPSTFKGTPTAENPEYIGLDVFPV 1253  
 DB 1197 LTPGGAGAPQHPHPPAPASPAFDNLVYWDQDPERGAPSTFKGTPTAENPEYIGLDVFPV 1255  
 RESULT 5  
 AAM92406  
 ID AAM92406 standard; Protein; 1255 AA.  
 AC AAM92406;  
 XX 21-APR-1999 (first entry)  
 DT 21-APR-1999 (first entry)  
 XX Human HER-2/neu oncogene protein.  
 DE Human HER-2/neu oncogene protein.  
 XX  
 KW HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;  
 KM malignancy; treatment; tumour.  
 XX Homo sapiens.  
 OS  
 XX  
 FH Key Location/Qualifiers  
 FT Region 676..1255  
 FT /note="region which elicits immune response"  
 XX  
 FN US869445-A.  
 XX 09-FEB-1999.  
 PD 09-FEB-1999.  
 XX  
 PF 01-APR-1996; 96US-0625101.  
 XX  
 PR 01-APR-1996; 96US-0625101.  
 PR 17-MAR-1993; 93US-0033644.  
 PR 12-AUG-1993; 93US-0106112.  
 PR 31-MAR-1995; 95US-0414417.  
 XX  
 PA (UNIM) UNIV WASHINGTON.  
 XX  
 PI Cheever MA, Disis ML;  
 XX WPI; 1999-152835/13.  
 DR N-PSDB; AAX01912.  
 XX  
 XX Use of HER-2/neu polypeptides - for eliciting an immune response to  
 PT an HER-2/neu associated malignancy, particularly for treating or  
 PT preventing tumours  
 XX  
 XX Claim 3; Column 31-38; 26pp; English.  
 XX  
 CC This sequence represents the human HER-2/neu oncogene protein. A fragment

CC of this protein is used in a method for eliciting or enhancing an immune  
 CC response to HER-2/neu protein. The polypeptide can stimulate T cells and  
 CC B cells to produce an immune response to the HER-2/neu protein. The  
 CC method can be used for immunisation against a malignancy in which the  
 CC HER-2/neu oncogene is associated and in the treatment of an existing  
 CC tumour, or to prevent tumour occurrence or reoccurrence.  
 CC  
 XX

SQ Sequence 1255 AA;  
 Query Match 96.9%; Score 6604; DB 20; Length 1255;  
 Best Local Similarity 96.7%; Pred. No. 0;  
 Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

QY 1 MELALACRGLLALALPGAASTQVCTGDMKLRLPASETHLDMRLHYOGQVQGNL 60  
 DB 1 MELALACRGLLALALPGAASTQVCTGDMKLRLPASETHLDMRLHYOGQVQGNL 60  
 QY 61 ELTYLPTNALSFLQDIQEVQGVYLIANQVQVPLQRLIRYRGTLFEDNYALVLDNG 120  
 DB 61 ELTYLPTNALSFLQDIQEVQGVYLIANQVQVPLQRLIRYRGTLFEDNYALVLDNG 120  
 QY 121 DPLNNTPTVAGSPGGLRELQRLSLTEILKGGVLIQRPOLCYODTITLWKDIFHKNOILA 180  
 DB 121 DPLNNTPTVAGSPGGLRELQRLSLTEILKGGVLIQRPOLCYODTITLWKDIFHKNOILA 180  
 QY 181 LTLIDNRSRACHPCSPMKCKGRCWGESSEDCQSLRTVYACGACAKCKPLPTDCHEQC 240  
 DB 181 LTLIDNRSRACHPCSPMKCKGRCWGESSEDCQSLRTVYACGACAKCKPLPTDCHEQC 240  
 QY 241 AAGCTGPKKSDCLACHFNHSGICEHCAALTYNPTFESMNPBGRTPGASCTACP 300  
 DB 241 AAGCTGPKKSDCLACHFNHSGICEHCAALTYNPTFESMNPBGRTPGASCTACP 300  
 QY 301 YNVLSTDVSGCTLVCPQLHNOEYTAEDGTORCEKCKSPKACVYGLQMOYIKANSXETGIT 360  
 DB 301 YNVLSTDVSGCTLVCPQLHNOEYTAEDGTORCEKCKSPKACVYGLQMOYIKANSXETGIT 360  
 QY 361 ELLEPACCKKIFGSLATLPESSFDGDPASNTAPLOPELOVFEETLEETGLYISAMPDSL 420  
 DB 361 IOEFACCKKIFGSLATLPESSFDGDPASNTAPLOPELOVFEETLEETGLYISAMPDSL 420  
 QY 421 DLSVFQNLQVIRGRIIHNAGVSLTLOGLISWGLRLSELGSLALIHNNHLCVHTV 480  
 DB 421 DLSVFQNLQVIRGRIIHNAGVSLTLOGLISWGLRLSELGSLALIHNNHLCVHTV 480  
 QY 481 PWDQLFRNPHQALLHTANPEDECVGEGLACHQLCARHGWGPGPTQCVNCSQFLRGEC 540  
 DB 481 PWDQLFRNPHQALLHTANPEDECVGEGLACHQLCARHGWGPGPTQCVNCSQFLRGEC 540  
 QY 541 VEECRVLOGLPREYVNAHRLCPHCECOPONGSVTCFGEADOCVACAHYKDPFCVAC 600  
 DB 541 VEECRVLOGLPREYVNAHRLCPHCECOPONGSVTCFGEADOCVACAHYKDPFCVAC 600  
 QY 601 PSGVXPDLSTYMPIMKFPDEGACOPCPINCTHSCVDLDKGPASQRPASPLTSISAVVG 660  
 DB 601 PSGVXPDLSTYMPIMKFPDEGACOPCPINCTHSCVDLDKGPASQRPASPLTSISAVVG 660  
 QY 661 ILLVYVLGVYRGILL-----FNNFTVSFWLAVPVKYSASHEPLTPSGAMPQAOQMRILK 714  
 DB 661 ILLVYVLGVYRGILL-----FNNFTVSFWLAVPVKYSASHEPLTPSGAMPQAOQMRILK 714  
 QY 715 ETELKRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIDELAVYVAG 774  
 DB 715 ETELKRVKVLGSGAFGVYKGIWIPDGENVKIPVAIKVLRNTSPKANKEIDELAVYVAG 774  
 QY 775 VGSFVYSRLGLICLSTVQLVTLQMLPYGGLDHTVENRGLASQDLNWKQIAGMSYL 834  
 DB 775 VGSFVYSRLGLICLSTVQLVTLQMLPYGGLDHTVENRGLASQDLNWKQIAGMSYL 834  
 QY 835 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGKVPIMKMALESILR 894  
 DB 837 EDVRLVHRDLAARNLVKSPNHVKITDGLARLLDIDETEHADGKVPIMKMALESILR 896

QY 895 RRFTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVIMIN 954  
 DB 897 RRFTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVIMIN 956  
 QY 955 KCMWIDSECRPRFRELVSFGRMARDPQRFVIVIONEDGPASPLDSTYRSLLDDMDGD 1014  
 DB 957 KCMWIDSECRPRFRELVSFGRMARDPQRFVIVIONEDGPASPLDSTYRSLLDDMDGD 1016  
 QY 1015 LVDAEYLYVPOGFFCPDPAPQAGMHHRRSSSTRSGGDLTLGLEPSEEARSPPLA 1074  
 DB 1017 LVDAEYLYVPOGFFCPDPAPQAGMHHRRSSSTRSGGDLTLGLEPSEEARSPPLA 1076  
 QY 1075 PSEGASDVFDGDLGMAKAGLQSLPTHDPSPLOKYSBDPTVPLPSENDGVAAVLTCSPQ 1134  
 DB 1077 PSEGASDVFDGDLGMAKAGLQSLPTHDPSPLOKYSBDPTVPLPSENDGVAAVLTCSPQ 1136  
 QY 1135 PEYVNPDPVRPQPSREGPLPAARPAAGATLERAKTTSFGKNGVVKDYFAFGAVENPEY 1194  
 DB 1137 PEYVNPDPVRPQPSREGPLPAARPAAGATLERAKTTSFGKNGVVKDYFAFGAVENPEY 1196  
 QY 1195 LTPGGGAAPQHPHAPAFSPFNDLYWMDQDPERGAPESTFKGTPTAENPEYLGDPV 1253  
 DB 1197 LTPGGGAAPQHPHAPAFSPFNDLYWMDQDPERGAPESTFKGTPTAENPEYLGDPV 1255

## RESULT 7

1D AAB21198 standard; protein; 1255 AA.

XX AAB21198;

DT 12-JAN-2001 (first entry)

DE Human HER-2/neu protein.

XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;

KM breast cancer; prostate cancer; ovarian cancer; lung cancer;

XX Homo sapiens.

XX MO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000MO-US02164.

XX 29-JAN-1999; 99US-0117976.

PA (CORI-) CORIXA CORP.

PA (SMK) SMITHKLINE BEECHAM.

PI Cheever MA, Gheysen D;

DR MPI: 2000-505976/45.

DR N-PSDB; AAB89736.

PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins

PT useful for vaccinating against breast, ovarian, colon, lung and

PT prostate cancers -

XX Claim 52; Fig 7; 128pp; English.

CC The present sequence is the human HER-2/neu protein. It is a member of  
 CC the tyrosine kinase family of receptor-like glycoproteins and shows  
 CC homology to the epidermal growth factor receptor (EGFR). It probably  
 CC plays a part in cell growth and/or differentiation. The HER-2/neu  
 CC gene is an oncogene. An HER-2/neu fusion protein comprising a  
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation  
 CC domain may be used to treat or prevent cancer by eliciting or  
 CC enhancing an immune response to the HER-2/neu protein. It may be used  
 CC to treat malignancies such as breast, ovarian, colon, lung and  
 CC prostate cancers, and may be used as an antigen to vaccinate against

CC these neoplasias.

XX Sequence 1255 AA;

Query Match 96.9%; Score 6604; DB 21; Length 1255;  
 Best Local Similarity 96.7%; Pred. No. 0;  
 Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

QY 1 MELALCRWGLLALLPPGASLQVCTGDMKRLPASPETHLDMRLHYOGQVQVQNL 60  
 DB 1 MELALCRWGLLALLPPGASLQVCTGDMKRLPASPETHLDMRLHYOGQVQVQNL 60  
 QY 61 ELTYLPTNASTSLQDIOEVQGVLLAHNQVQVPLQRRLIRVGTQLFEDNYALAVLNG 120  
 DB 61 ELTYLPTNASTSLQDIOEVQGVLLAHNQVQVPLQRRLIRVGTQLFEDNYALAVLNG 120  
 QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILLKGVLLQRPOLCYQDTILLKDIFFHKNOLA 180  
 DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILLKGVLLQRPOLCYQDTILLKDIFFHKNOLA 180  
 QY 181 LTLIDNRSRACHPCSPMCKGRCKWESSEDCQSLRTVCAQGCARCKPLPTDCCHEQC 240  
 DB 181 LTLIDNRSRACHPCSPMCKGRCKWESSEDCQSLRTVCAQGCARCKPLPTDCCHEQC 240  
 QY 181 LTLIDNRSRACHPCSPMCKGRCKWESSEDCQSLRTVCAQGCARCKPLPTDCCHEQC 240  
 DB 181 LTLIDNRSRACHPCSPMCKGRCKWESSEDCQSLRTVCAQGCARCKPLPTDCCHEQC 240  
 QY 241 AACCTGPKASDCLACLHFNHSGICELHCPALVYNTDTESMPNPEGRYTFGASCYTACP 300  
 DB 241 AACCTGPKASDCLACLHFNHSGICELHCPALVYNTDTESMPNPEGRYTFGASCYTACP 300  
 QY 301 YNTLSTDVSGCTLVCELNQVTAEDGTQCEKSPCARVCYGLCMQYIKANSKFIKIT 360  
 DB 301 YNTLSTDVSGCTLVCELNQVTAEDGTQCEKSPCARVCYGLCMQYIKANSKFIKIT 360  
 QY 301 YNTLSTDVSGCTLVCELNQVTAEDGTQCEKSPCARVCYGLCMQYIKANSKFIKIT 360  
 DB 301 YNTLSTDVSGCTLVCELNQVTAEDGTQCEKSPCARVCYGLCMQYIKANSKFIKIT 360  
 QY 361 ELEFAGCKKIFGSLAFIPESFDDPASNTAPLOEOLQVETLEITGLYISAMPDSL 420  
 DB 361 IOEFAGCKKIFGSLAFIPESFDDPASNTAPLOEOLQVETLEITGLYISAMPDSL 420  
 QY 421 DLSVFNQVYIRRIHNGAYSLTLOGLSIWLGLSLRLSGLLIHNHNLCEVHTV 480  
 DB 421 DLSVFNQVYIRRIHNGAYSLTLOGLSIWLGLSLRLSGLLIHNHNLCEVHTV 480  
 QY 481 PMDQLFNPQALHTANRPEDECVBEGLAHQULCARHGWGFPQVNCOSQFLGOEC 540  
 DB 481 PMDQLFNPQALHTANRPEDECVBEGLAHQULCARHGWGFPQVNCOSQFLGOEC 540  
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFPPADQCVACAHYKDPFCVARC 600  
 DB 541 VEECRVLOGLPREYVNAHCLPCHPECOPOGNSVTCFPPADQCVACAHYKDPFCVARC 600  
 QY 601 PSGVKPDLSTYMPIMKPEDEGACQPCGCTINCTHSCVDLDDKCPAEOQASPLTSIISAVVG 660  
 DB 601 PSGVKPDLSTYMPIMKPEDEGACQPCGCTINCTHSCVDLDDKCPAEOQASPLTSIISAVVG 660  
 QY 661 ILVVLGVVFFGLI-----FNNFTVSEFLRVKVSASHLBEPLTSGAMPQAOQRILK 714  
 DB 661 ILVVLGVVFFGLI-----FNNFTVSEFLRVKVSASHLBEPLTSGAMPQAOQRILK 714  
 QY 715 ETELARKYKVLGSGAFGVYKGIWIPGSENKIVALKVLRNTSPANKKILDEAYVMAG 774  
 DB 715 ETELARKYKVLGSGAFGVYKGIWIPGSENKIVALKVLRNTSPANKKILDEAYVMAG 774  
 QY 717 ETELARKYKVLGSGAFGVYKGIWIPGSENKIVALKVLRNTSPANKKILDEAYVMAG 776  
 DB 717 ETELARKYKVLGSGAFGVYKGIWIPGSENKIVALKVLRNTSPANKKILDEAYVMAG 776  
 QY 775 VGSPPVSRLLGICLTSTVQLVLTQMPGCLLDHVRNKGRLSGODLLNMCMQIAKMSYL 834  
 DB 775 VGSPPVSRLLGICLTSTVQLVLTQMPGCLLDHVRNKGRLSGODLLNMCMQIAKMSYL 834  
 QY 835 EDVRLVARDLAANNVLKSPNNHYKITDPGLARLLIDETRYHADGKVPKIMVALSILR 894  
 DB 835 EDVRLVARDLAANNVLKSPNNHYKITDPGLARLLIDETRYHADGKVPKIMVALSILR 894  
 QY 895 RRFTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVIMIN 954  
 DB 897 RRFTHQSDVMSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPPICITIDVIMIN 956  
 QY 955 KCMWIDSECRPRFRELVSFGRMARDPQRFVIVIONEDGPASPLDSTYRSLLDDMDGD 1014

Db 957 KCMWIDSECRPRFELVSEFSRMARDPQRFVIONEDLGASPLDSTFYSLLEDDMGD 1016  
 Qy 1015 LVDAEYLVPQOGFFCDDPAPAGAGMVRHRRSSSTSGGGDLTLGLEPSEBAPRSPLA 1074  
 Db 1017 LVDAEYLVPQOGFFCDDPAPAGAGMVRHRRSSSTSGGGDLTLGLEPSEBAPRSPLA 1076  
 Qy 1075 PSEBAGSDVDGDLGMAAGKLGSLPTHDSPLQRYSEDPVLPSESTDGYVAPLTCSPQ 1134  
 Db 1077 PSEBAGSDVDGDLGMAAGKLGSLPTHDSPLQRYSEDPVLPSESTDGYVAPLTCSPQ 1136  
 Qy 1135 PEYVQPDVPRPPSPREGLPAARPAAGATLERPKTSLPGKNGVYKDVAFGAVENPEY 1194  
 Db 1137 PEYVQPDVPRPPSPREGLPAARPAAGATLERPKTSLPGKNGVYKDVAFGAVENPEY 1196  
 Qy 1195 LTPGGAPQHPHPAPSPAFDNLVYWDQDPREBAPSPFTFKGPTAENPEYLGLDV 1253  
 Db 1197 LTPGGAPQHPHPAPSPAFDNLVYWDQDPREBAPSPFTFKGPTAENPEYLGLDV 1255

RESULT 8  
 AAY84780  
 ID AAY84780 standard; Protein; 1255 AA.

XX AAY84780;  
 AC  
 XX 08-AUG-2000 (first entry)  
 DT  
 XX

DE Amino acid sequence of the SPICE erbB-2 receptor protein.

XX SPICE erbB-2 receptor protein; cell transformation disorder; cancer;  
 KM tumor cell proliferation; tissue degeneration; arthropathy;  
 KM bone resorption; inflammatory disease; degenerative disorder;  
 KM wound healing.

OS Homo sapiens.

XX MO200020579-A1.

XX 13-APR-2000.

XX 01-OCT-1999; 99MO-CA00912.

XX 02-OCT-1998; 98US-0165192.

XX (UYMC-) UNIV MCMASTER.

XX Muller WJ, Siegel PM;

XX WPI; 2000-303768/26.

XX N-PSDB; AAA14812.

XX Nucleic acid encoding an erbB 2 receptor protein designated SPICE

XX erbB-2, inhibitors of the protein are useful for treatment of cancer -

XX Claim 3; Fig 2; 60pp; English.

XX The present sequence represents a SPICE erbB-2 receptor protein. The  
 CC protein has an in-frame deletion of 16 amino acids, 2 of which are  
 CC conserved cysteine residues, compared to the unspliced protein. The  
 CC erbB-2 polynucleotide is used to construct probes for detecting  
 CC disorders of cell transformation such as cancer. Antibodies to the  
 CC protein may be used to detect SPICE erbB-2 in a sample. Agents  
 CC (e.g. antisense oligonucleotides) which inhibit the expression of  
 CC SPICE erbB-2 are useful for reducing tumor cell proliferation and  
 CC treating cancer. Substances which stimulate SPICE erbB-2 are useful  
 CC for treating conditions involving damaged cells including conditions  
 CC in which degeneration of tissue occurs, such as arthropathy, bone  
 CC resorption, inflammatory diseases, degenerative disorders of the  
 CC central nervous system and wound healing.

XX Sequence 1255 AA;

Query Match 96.9%; Score 6604; DB 21; Length 1255;  
 Best Local Similarity 96.7%; Pred. No. 0;  
 Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

Qy 1 MELAALCRWGLLJALLPPGAATVCTGTDMKRLPASPETHLDMRLHYOCQVYVQNTL 60  
 Db 1 MELAALCRWGLLJALLPPGAATVCTGTDMKRLPASPETHLDMRLHYOCQVYVQNTL 60  
 Qy 61 ETLVLPYTNASISFLDDIQEVQGVYLIANQVQVLPQRLIRYRGQLFEDNALAVLDNG 120  
 Db 61 ETLVLPYTNASISFLDDIQEVQGVYLIANQVQVLPQRLIRYRGQLFEDNALAVLDNG 120  
 Qy 121 DPLNNTTPVTGASPGGLBELQRLSTELIKGVLIRNPOLCYQDTILMKDIFHNQOLA 180  
 Db 121 DPLNNTTPVTGASPGGLBELQRLSTELIKGVLIRNPOLCYQDTILMKDIFHNQOLA 180  
 Qy 181 LTLIDTNRRAACHPCSPKCKSGRCESESDCOSLTRIVCAGGACARCKPLPTDCHEQC 240  
 Db 181 LTLIDTNRRAACHPCSPKCKSGRCESESDCOSLTRIVCAGGACARCKPLPTDCHEQC 240  
 Qy 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRTFGASCYTACP 300  
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPREGRTFGASCYTACP 300  
 Qy 301 YNYLSTDVSGCTLVCPLNQVETADGTQRCCKSPCARVCYGLGMQYIKANSKFTIGT 360  
 Db 301 YNYLSTDVSGCTLVCPLNQVETADGTQRCCKSPCARVCYGLGMQYIKANSKFTIGT 360  
 Qy 361 ELLEFAGCKKIFESLAFLESFPGDPASTAPLQEPQLQVFELEITGLYISAMPDSIP 420  
 Db 361 IGEFAGCKKIFESLAFLESFPGDPASTAPLQEPQLQVFELEITGLYISAMPDSIP 420  
 Qy 421 DLSVFQNLQVIRGILHNGAVSLTLQGLISLGLRSRELQSGALILHNHILCFVHTV 480  
 Db 421 DLSVFQNLQVIRGILHNGAVSLTLQGLISLGLRSRELQSGALILHNHILCFVHTV 480  
 Qy 481 PWDQLFRNPQALLHTANRBEDECYGEGLACHQLCARHGWCPGPTQCVNCSQIFRGQC 540  
 Db 481 PWDQLFRNPQALLHTANRBEDECYGEGLACHQLCARHGWCPGPTQCVNCSQIFRGQC 540  
 Qy 541 VECCRYLQGLPREYVYNAHCLPCHBECPQONOSVTCFGEADQCAACHYKDPFCVABC 600  
 Db 541 VECCRYLQGLPREYVYNAHCLPCHBECPQONOSVTCFGEADQCAACHYKDPFCVABC 600  
 Qy 601 PGGVPRDLSYMPIMKFPDEBACQPCPINCSTHSCVDLDDKCGPABQRASPLTISIAYVG 660  
 Db 601 PGGVPRDLSYMPIMKFPDEBACQPCPINCSTHSCVDLDDKCGPABQRASPLTISIAYVG 660  
 Qy 661 ILVVVLGVVFGLI-----FNNFTVSFMLRVPKVSASHLEPLTPSGAMPQAOQMRILK 714  
 Db 661 ILVVVLGVVFGLI-----FNNFTVSFMLRVPKVSASHLEPLTPSGAMPQAOQMRILK 714  
 Qy 715 ETELKRVKVLGSGAGTAYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVMAG 774  
 Db 715 ETELKRVKVLGSGAGTAYKGIWIPDGENVKIPVAIKYLRNTSPKANKELIDEAYVMAG 774  
 Qy 775 VSSPVVSRLLGICLSTYQVLTQMLPYGCLLDHVENRGLSGQDLVWCMQIAKMSYL 834  
 Db 775 VSSPVVSRLLGICLSTYQVLTQMLPYGCLLDHVENRGLSGQDLVWCMQIAKMSYL 834  
 Qy 835 EDVRLVHMDLAARNVLYVSPNHVKITDGLARLIDIDTEVYADGKPIKMALESTLR 894  
 Db 835 EDVRLVHMDLAARNVLYVSPNHVKITDGLARLIDIDTEVYADGKPIKMALESTLR 894  
 Qy 897 RFTHQSDVWSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPQPICTIDVYMTV 954  
 Db 897 RFTHQSDVWSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPQPICTIDVYMTV 954  
 Qy 955 KCMWIDSECRPRFELVSEFSRMARDPQRFVIONEDLGASPLDSTFYSLLEDDMGD 1014  
 Db 955 KCMWIDSECRPRFELVSEFSRMARDPQRFVIONEDLGASPLDSTFYSLLEDDMGD 1014  
 Qy 1015 LVDAEYLVPQOGFFCDDPAPAGAGMVRHRRSSSTSGGGDLTLGLEPSEBAPRSPLA 1074

DB 1017 LVDAEYLVPQGFCCDPAPAGAGWVHHRSSSTRSGGDLTLGLEPSEEEAPRSPLA 1076  
 QY 1075 PSEAGSDVFDGDLGMAKAAGLQSLPTHDPSPLOQRYSEDPVLPSESTDGVAPLTCSPQ 1134  
 DB 1077 PSEAGSDVFDGDLGMAKAAGLQSLPTHDPSPLOQRYSEDPVLPSESTDGVAPLTCSPQ 1136  
 QY 1135 PEYVNOPDVRPQPSPREGGLPAARPAAGATLERAKTILSPGKNGVYKQVFAFGAVENPEY 1194  
 DB 1137 PEYVNOPDVRPQPSPREGGLPAARPAAGATLERAKTILSPGKNGVYKQVFAFGAVENPEY 1196  
 QY 1195 LTPQGAAPQHPAPAPAPADNLYYMDODPPEAGAPSPFPGTPTAENPEYLGLDVPPV 1253  
 DB 1197 LTPQGAAPQHPAPAPAPADNLYYMDODPPEAGAPSPFPGTPTAENPEYLGLDVPPV 1255

## RESULT 9

AAB85458 ID AAB85458 standard; Protein; 1255 AA.

AC AAB85458;

DT 25-SEP-2001 (first entry)

DE Human HER-2/neu protein.

KM Antigen-presenting cell; immunogenic; immune response: HER-2/neu;

KM oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.

OS Homo sapiens.

PN WO200153463-A2.

PD 26-JUL-2001.

PF 19-JAN-2001; 2001WO-US01850.

PR 21-JAN-2000; 2000US-0177545.

PA (CORI-) CORIXA CORP.

PI Cheever MA, Hand-Zimmermann S;

DR WPI; 2001-476112/51.

DR N-PSDB; AAH23392.

PT New antigen-presenting cells, useful as vaccines for eliciting or

PT enhancing an immune response to HER-2/neu protein, particularly useful

PT for treating or preventing cancer, e.g. breast cancer

PS Claim 2; Page 41-46; 49pp; English.

CC The invention provides an isolated antigen-presenting cell, which

CC expresses at least an immunogenic portion of a polypeptide that produces

CC an immune response to HER-2/neu protein. The antigen-presenting cells are

CC useful as vaccines for eliciting or enhancing an immune response to

CC HER-2/neu protein, particularly in treating or preventing malignancies in

CC which the HER-2/neu oncogene is associated. Specifically, these are

CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,

CC colon, lung or prostate cancers. The present sequence represents

CC the human HER-2/neu protein (also known as p185 or c-erbB2).

SO Sequence 1255 AA;

Query Match 96.9%; Score 6604; DB 22; Length 1255;

Best Local Similarity 96.7%; Pred. No. 0;

Matches 1318; Conservative 11; Mismatches 20; Indels 10; Gaps 2;

QY 1 MELALCRWGLLALPPGAASSTQVCTGDMKLRPAAPETHLDMKLRHYOGCCVYQGNL 60

DB 1 MELALCRWGLLALPPGAASSTQVCTGDMKLRPAAPETHLDMKLRHYOGCCVYQGNL 60

QY 61 ELTYLPTNASLSTFLQDIOEVQGYLIAHNOVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120

DB 61 ELTYLPTNASLSTFLQDIOEVQGYLIAHNOVRQVPLQRLRIVRGTLFEDNYALAVLDNG 120  
 QY 121 DPLANTTPYTGASPGGLARELQRLSTLEILKGVLLIQRNEPOLCYDTIIMKDI FHKNNQLA 180  
 DB 122 DPLANTTPYTGASPGGLARELQRLSTLEILKGVLLIQRNEPOLCYDTIIMKDI FHKNNQLA 180  
 QY 181 LTLIDNRBRACHPGSPMKCKSRMGESSEDCQSLTRYTCAGGACRCKRPLTDCCHQC 240  
 DB 181 LTLIDNRBRACHPGSPMKCKSRMGESSEDCQSLTRYTCAGGACRCKRPLTDCCHQC 240  
 QY 241 AAGCTPKASDCLACHFNHSGICEILHCPALVTYNTDTPESMPNEGRYTFGASCTYACP 300  
 DB 241 AAGCTPKASDCLACHFNHSGICEILHCPALVTYNTDTPESMPNEGRYTFGASCTYACP 300  
 QY 301 YNYLSTDVSGCTLVCPLEHNOEYTABEDGTQRCCKSPKPCARVCYGLMOYIKANSKTI 360  
 DB 301 YNYLSTDVSGCTLVCPLEHNOEYTABEDGTQRCCKSPKPCARVCYGLMOYIKANSKTI 360  
 QY 361 ELFFACCKXIFGSLAFPESPFGDDPASNTAPLOPELOQVFEETLEETGYLYISAMPD 420  
 DB 361 ELFFACCKXIFGSLAFPESPFGDDPASNTAPLOPELOQVFEETLEETGYLYISAMPD 420  
 QY 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGISMLGRSLRELSSGLALIHNNTHLCVHTV 480  
 DB 421 DLSVFQNLQVIRGRILHNGAYSLLTQGLGISMLGRSLRELSSGLALIHNNTHLCVHTV 480  
 QY 481 PMQULFRNPHQALHNTANPEDECVGEGALACHQOLCARGHCMKPGPPOCVNCSQFLAGQ 540  
 DB 481 PMQULFRNPHQALHNTANPEDECVGEGALACHQOLCARGHCMKPGPPOCVNCSQFLAGQ 540  
 QY 541 VEECRVLQGLPREYVNAARHCLPCHPEQOPONGSVTCFGEADQCVACAHYKDPFCVAC 600  
 DB 541 VEECRVLQGLPREYVNAARHCLPCHPEQOPONGSVTCFGEADQCVACAHYKDPFCVAC 600  
 QY 601 PGGVXPDLSTYMPWKPPDEEGACQCPINCTHSQVLDLQKGAERASPLSTVAVVG 660  
 DB 601 PGGVXPDLSTYMPWKPPDEEGACQCPINCTHSQVLDLQKGAERASPLSTVAVVG 660  
 QY 661 ILVVLGVVFGILIKRQOKIRKYMRLLOTEL----VEPLTSGAMPNQAQRIILK 714  
 DB 661 ILVVLGVVFGILIKRQOKIRKYMRLLOTEL----VEPLTSGAMPNQAQRIILK 714  
 QY 715 ETELKRVKVLGSGAFSTYKGIWIPDGENVKIPVAIKVIRENTSPRANKIIDEAYVAG 774  
 DB 715 ETELKRVKVLGSGAFSTYKGIWIPDGENVKIPVAIKVIRENTSPRANKIIDEAYVAG 774  
 QY 775 VGSPPVSRLLGLTSTVQLVQLMFYGCLDHYENRGRIGSODLLNMCQIAKMSYL 834  
 DB 775 VGSPPVSRLLGLTSTVQLVQLMFYGCLDHYENRGRIGSODLLNMCQIAKMSYL 834  
 QY 835 EDVRLVHRDLAARNVLVKSBNHYKIDFGIARLLIDETRYHADGGKVPIMMALESILR 894  
 DB 835 EDVRLVHRDLAARNVLVKSBNHYKIDFGIARLLIDETRYHADGGKVPIMMALESILR 894  
 QY 895 RRTTHGSDWVSQVYTWELMTFGAKYDGIIPAEIIDLLEKGRLLQOPRCTIDVYMIW 954  
 DB 895 RRTTHGSDWVSQVYTWELMTFGAKYDGIIPAEIIDLLEKGRLLQOPRCTIDVYMIW 954  
 QY 955 KCMWISSECRPRELVESESRMARPORFVVIQNEDLGPAASLDSTFYRSLLEDDMGD 1014  
 DB 955 KCMWISSECRPRELVESESRMARPORFVVIQNEDLGPAASLDSTFYRSLLEDDMGD 1014  
 QY 1015 LVDAEYLVPQGFCCDPAPAGAGWVHHRSSSTRSGGDLTLGLEPSEEEAPRSPLA 1074  
 DB 1015 LVDAEYLVPQGFCCDPAPAGAGWVHHRSSSTRSGGDLTLGLEPSEEEAPRSPLA 1074  
 QY 1075 PSEAGSDVFDGDLGMAKAAGLQSLPTHDPSPLOQRYSEDPVLPSESTDGVAPLTCSPQ 1134  
 DB 1075 PSEAGSDVFDGDLGMAKAAGLQSLPTHDPSPLOQRYSEDPVLPSESTDGVAPLTCSPQ 1134  
 QY 1135 PEYVNOPDVRPQPSPREGGLPAARPAAGATLERAKTILSPGKNGVYKQVFAFGAVENPEY 1194  
 DB 1135 PEYVNOPDVRPQPSPREGGLPAARPAAGATLERAKTILSPGKNGVYKQVFAFGAVENPEY 1194

Db 1137 PEYNNQPVRRQQRPSREGGRLPAARPAACATLERPTLLSGKNGVXVDYPAFGAGVENDEY 1196

QY 1195 LTPOGGAAPQHPPEAFSPAFDNLVYMQDPPERGAPSTFKGTPTAENPEYLGLDVPY 1253

Db 1197 LTPOGGAAPQGHPRPAFSPAFDNLVYMQDPPERGAPSTFKGTPTAENPEYLGLDVPY 1255

RESULT 10  
AAG88267  
ID AAG88267 standard; Protein; 1255 AA

AC AAG88267;

DT 11-SEP-2001 (first entry)

HER2/neu amino acid sequence

KM Human: HBR2/neu: vaccine: human leukocyte antigen; HLA; T cell;  
 KM immune response; epitope: tumour; cancer: cytotoxic; immunostimulant  
 KM tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.  
 XX  
 OS Homo sapiens.

PN WO200141787-A1

PD 14-JUN-2001

PF 11-DEC-2000; 2000WO-US33591

PR 10-DEC-1999; 99US-0458299

PA (EPIM-) EPIMUNE INC

PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E,

DR WPI; 2001-374995/39

PT An isolated prepared HER2/new epitope useful in a vaccine for inducing  
PT cellular immune responses for the prevention and treatment of cancer  
XX  
ES Disclosure, Page 15, 199pp; English.

The present invention describes isolated prepared HER2/neu epitopes (1). Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is culture in vitro and binds to a complex of an epitope (I), bound to a human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I) and a second epitope and the peptide is less than 50 contiguous amino acids that have 100% identity with a native peptide sequence of HER2/neu; (3) a vaccine composition (III) comprising (II) and a pharmaceutical excipient; (4) an isolated nucleic acid encoding a peptide comprising (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic and immunostimulant activities, and can be used in vaccines. (I), (II) and (III) are useful for inducing cellular immune responses for the prevention and treatment of cancer. (I) and (II) are useful for monitoring or evaluating an immune response to a tumour-associated antigen when incubated with a T lymphocyte sample from a patient and detecting the presence of bound T lymphocyte to (I) or (II). Epitope based vaccines mean that immunosuppressive epitopes that may be present in whole antigens may be avoided. Selected epitopes may be combined to enhance immunogenicity. The possible pathological side effects caused by infectious agents or whole protein antigen is eliminated. The vaccine provides the ability to direct and focus an immune response to multiple selected antigens from the same pathogen. Epitope-based anti-tumour vaccines provides the opportunity to combine epitopes derived from multiple tumour-associated molecules addressing the problem of tumour-tumour variability and reducing the likelihood of tumour escape due to antigen loss. AAG8266 to AAG89121 represent amino acid sequences used in the exemplification of the present invention.

Sequence 1255 AA;

Query Match	96.9%; Score 6604; DB 22; Length 1255
-------------	---------------------------------------

Best Local Similarity 96.7%; Pred. No. 0;  
Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2

QY	1	MELAAORMGILLALLPPGAASOVCTDMKLELPASPETHMLRLUYOOCVAVQNL	60
Db	1	MELAAORMGILLALLPPGAASOVCTGTDKMLPASPETHMLRLUYOOCVAVQNL	60
QY	61	ELTYLPTNASLSPQDIOEVGYVLIANNOYVPLQRLRIVRGTOLEEDNYALAVDNG	120
Db	61	ELTYLPTNASLSPQDIOEVGYVLIANNOYVPLQRLRIVRGTOLEEDNYALAVDNG	120
QY	121	DLNNNTPYVGASGGJRELQBSLETELKSGVLIQNNPOCXYDDITLMKOIPIHNNOLA	180
Db	121	DLNNNTPYVGASGGJRELQBSLETELKSGVLIQNNPOCXYDDITLMKOIPIHNNOLA	180
QY	181	LTLIDTRNSRACHPCSPMCKSGRCWGSSEEDCOGLTRVCAGGACRCKGRLPTDCHEQC	240
Db	181	LTLIDTRNSRACHPCSPMCKSGRCWGSSEEDCOGLTRVCAGGACRCKGRLPTDCHEQC	240
QY	241	AAGCTGPKHSCCLCHFPNHSIGTELMCPALVYNTDTFESMPNPEGVYTTGASCVTACP	300
Db	241	AAGCTGPKHSCCLCHFPNHSIGTELMCPALVYNTDTFESMPNPEGVYTTGASCVTACP	300
QY	301	YNYSTDVGSCTVLCPLHNOEVAEDGTORCEKSKPCARCYGAGMOYIANSKFIQT	360
Db	301	YNYSTDVGSCTVLCPLHNOEVAEDGTORCEKSKPCARCYGAGMHLEEVRAVTSAN	360
QY	361	ELBPAQCKKIFGSLAPLPESFDGDPASNTABLOPELOQVFTTEITGYLYISAMPDLP	420
Db	361	IQEPAQCKKIFGSLAPLPESFDGDPASNTABLOPELOQVFTTEITGYLYISAMPDLP	420
QY	421	DLSPFONLOVRGILHNGAYSLTLOGIGISWGLRSLRELGSGLALIHNNTHLCFVNTV	480
Db	421	DLSPFONLOVRGILHNGAYSLTLOGIGISWGLRSLRELGSGLALIHNNTHLCFVNTV	480
QY	481	FMQDLFRNPQALLHTANRPEDECVGEGLAHQLOCARHCWGPPTQCVNCSQFLRGEC	540
Db	481	FMQDLFRNPQALLHTANRPEDECVGEGLAHQLOCARHCWGPPTQCVNCSQFLRGEC	540
QY	541	VEBQRYVIOGPREVYNARHCLPCHPEOPQONGSTTCGRPADQCVACHYKDPPEVCARC	600
Db	541	VEBQRYVIOGPREVYNARHCLPCHPEOPQONGSTTCGRPADQCVACHYKDPPEVCARC	600
QY	601	PSGVKPDLSYMPIMKFPDEBEGACQPCFINCHSCVDDLDDKCCAPAEORASPLTISAVVG	660
Db	601	PSGVKPDLSYMPIMKFPDEBEGACQPCFINCHSCVDDLDDKCCAPAEORASPLTISAVVG	660
QY	661	ILLYVYVGVYFGIL-----FNNFYTSFWLRVYKVSASHLEPLTPSGAMNQAOMRLK	714
Db	661	ILLYVYVGVYFGILIKKROOKIRKTYTNRLLQETEL---VEPLTPSGAMNQAOMRLK	716
QY	715	ETEARKVYVGSAGFGYVYKGIWIPDGNVYIPIAIKYLRENTSPKANKETLDEAYYWAG	774
Db	717	ETEARKVYVGSAGFGYVYKGIWIPDGENYKIPAIKYLRENTSPKANKELLDENAYYWAG	776
QY	775	VGSPPYVBRLLIGTCLTSYVQVLTQMLPYGCLLDHYRENRGRUGSODLLNMCQIAKAGSYL	834
Db	777	VGSPPYVBRLLIGTCLTSYVQVLTQMLPYGCLLDHYRENRGRUGSODLLNMCQIAKAGSYL	836
QY	835	EDVAVLYRDLAARVNLVKSPPHVYKITPFGALRLDIDETEHAAOGKVPIMMALESTLR	894
Db	837	EDVAVLYRDLAARVNLVKSPPHVYKITPFGALRLDIDETEHAAOGKVPIMMALESTLR	896
QY	895	RRFTHQSDVMSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPOPPICITDVMYIMV	954
Db	897	RRFTHQSDVMSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPOPPICITDVMYIMV	956
QY	955	KCMWIDECRRPFEELVSEPSRMDRQRFVIONEDLGSPASPLDSTFYRLLDDDDMG	1014
Db	957	KCMWIDECRRPFEELVSEPSRMDRQRFVIONEDLGSPASPLDSTFYRLLDDDDMG	1016
QY	1015	LVDAEELVLPQGFCCDPAFGAAGMHHRRSSSTRSGGDDLTLGLIEPSEEAAPRSLA	1074



Db 1017 LVDAEEYLVPOQGFPCDPAAGAGMWHHRSSSTRSGGDLTLGLPSEEEAPRSPPLA 1076  
 Qy 1075 PSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOKYSDDPTVPLPSETDGVAPLTCSPQ 1134  
 Db 1077 PSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOKYSDDPTVPLPSETDGVAPLTCSPQ 1136  
 Qy 1135 PEYVNOQDVPRPQPSRREGPLPAARPAATLIERAKTILSPKNGVYKXVFAFGAVENPEY 1194  
 Db 1137 PEYVNOQDVPRPQPSRREGPLPAARPAATLIERAKTILSPKNGVYKXVFAFGAVENPEY 1196  
 Qy 1195 LTPOGGAPOPHPAPFSPAFDNLVYWDQDPPERGAPESTFKGTPTAENPEYLGLDVAV 1253  
 Db 1197 LTPOGGAPOPHPAPFSPAFDNLVYWDQDPPERGAPESTFKGTPTAENPEYLGLDVAV 1255

## RESULT 11

AAE24067  
 ID AAE24067 standard; Protein; 1255 AA.

AAE24067;

23-SEP-2002 (first entry)

Human Her-2 protein.

Human; Her-2; epidermal growth factor receptor 2; infection; cancer;

hyperproliferative disorder; prophylaxis; inflammation; antisense;

tumour; gene therapy; phosphothioate backbone.

Homo sapiens.

WO200222636-A1.

21-MAR-2002.

12-SEP-2001; 2001WO-US28572.

15-SEP-2000; 2000US-0663834.

(ISIS-) ISIS PHARM INC.

Bennett CF, Cowase LM;

WPI; 2002-471192/50.

DR N-PSDB; AAD38904.

Novel antisense oligonucleotide which modulates the expression of Human

Epidermal Growth Factor receptor, Her2, is useful for treating tumors

inflammation or to prevent infection in humans -

Example 13; Page 95-107; 116pp; English.

The invention relates to antisense compounds targeted to a nucleic

acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)

CC specifically hybridises with and inhibits the expression of Her2.

CC Antisense compounds of the invention are used for treating diseases or

CC conditions associated with Her2 such as hyperproliferative disorders

CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,

CC neutral or cardiac cancer. They are also useful prophylactically e.g.

CC to prevent or delay infection, inflammation and tumour formation. The

CC invention is also used in gene therapy. The present sequence is human

CC Her-2 protein.

CC

CC

Qy 61 ELTYLPTNASTSLTODICEVGVYLIHNOVQVPLQRLRIYRGTLFEEDNALAVLDNG 120  
 Db 61 ELTYLPTNASTSLTODICEVGVYLIHNOVQVPLQRLRIYRGTLFEEDNALAVLDNG 120  
 Qy 121 DPLANTTPTVAGSPGGLARELOLSLTELKGLVLIQORNQOLCYODITLWKDIFFKNNOLA 180  
 Db 121 DPLANTTPTVAGSPGGLARELOLSLTELKGLVLIQORNQOLCYODITLWKDIFFKNNOLA 180  
 Qy 181 LTLIDNRSBACHPGSPCKSGSCWCESSDQSLTRYCAGGACAKGRLPTDCHCQC 240  
 Db 181 LTLIDNRSBACHPGSPCKSGSCWCESSDQSLTRYCAGGACAKGRLPTDCHCQC 240  
 Qy 241 AAGCTGPKSDCLACLFHNSGICELHCAVLTNTDTESMNPREGRTFGASCYTAC 300  
 Db 241 AAGCTGPKSDCLACLFHNSGICELHCAVLTNTDTESMNPREGRTFGASCYTAC 300  
 Qy 301 YNYLSTDVSGCTLVCEPLHNOVTAEDGTORCEKSPCARVCGYGLMEHLREYRAVTSAN 360  
 Db 301 YNYLSTDVSGCTLVCEPLHNOVTAEDGTORCEKSPCARVCGYGLMEHLREYRAVTSAN 360  
 Qy 361 ELFEAGCKXIFGSLAFPLSPFODGPAASNTAPLOPELOVETLEETLGYLYISAMPDSL 420  
 Db 361 ELFEAGCKXIFGSLAFPLSPFODGPAASNTAPLOPELOVETLEETLGYLYISAMPDSL 420  
 Qy 421 DLSEFQNLQVTRGRILHNGAYSLTLQGLISWGLRSLBELSGSLA.IHNTLCEVHTV 480  
 Db 421 DLSEFQNLQVTRGRILHNGAYSLTLQGLISWGLRSLBELSGSLA.IHNTLCEVHTV 480  
 Qy 481 PMQULFRNPQALLHNTANPEDECEVEGLACHOLCARGHCMGPPQVCNCSQFLGQEC 540  
 Db 481 PMQULFRNPQALLHNTANPEDECEVEGLACHOLCARGHCMGPPQVCNCSQFLGQEC 540  
 Qy 541 VESCRVYQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600  
 Db 541 VESCRVYQGLPREYVNAHCLPCHPECPONGSVTCFGEADQCVACAHYKDPFCVARC 600  
 Qy 601 PSQVYKPLSTWPIWKPPDEEGACQPCPINCSTHCVLDLDDKCAEORASFLTSVAVVG 660  
 Db 601 PSQVYKPLSTWPIWKPPDEEGACQPCPINCSTHCVLDLDDKCAEORASFLTSVAVVG 660  
 Qy 661 ILVYVGVVFGILIKRQOKIRKYMRLDQTEL----VEPLTSGAMPNQAQRILK 714  
 Db 661 ILVYVGVVFGILIKRQOKIRKYMRLDQTEL----VEPLTSGAMPNQAQRILK 714  
 Qy 715 ETELKRVKVLGSGAFGVYKGIWIPGENYKIPVAIKVRENTSPANKELDEAYVAG 774  
 Db 715 ETELKRVKVLGSGAFGVYKGIWIPGENYKIPVAIKVRENTSPANKELDEAYVAG 774  
 Qy 775 VGSPIYVSRLLGICLTSTVQVLTQLMFYGCLLDHVRNRGBLGSODLLNMCMQIAKMSYL 834  
 Db 775 VGSPIYVSRLLGICLTSTVQVLTQLMFYGCLLDHVRNRGBLGSODLLNMCMQIAKMSYL 834  
 Qy 835 EDVRLVARDLAANVVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKXMALESILR 894  
 Db 835 EDVRLVARDLAANVVLKSPNHVKITDFGLARLLDIDETEHADGGKVPKXMALESILR 894  
 Qy 895 RRTFHOSDWSYGTWELMTFGAKYDGIIPAEIPDLKSGRLQOPPCITDVMYIW 954  
 Db 895 RRTFHOSDWSYGTWELMTFGAKYDGIIPAEIPDLKSGRLQOPPCITDVMYIW 954  
 Qy 955 KCMWIDECRPRELVESESRMAPQRFVYQNEDLGASPLDSTFYSLIEDDMGD 1014  
 Db 955 KCMWIDECRPRELVESESRMAPQRFVYQNEDLGASPLDSTFYSLIEDDMGD 1014  
 Qy 1015 LVDAEEYLVPOQGFPCDPAAGAGMWHHRSSSTRSGGDLTLGLPSEEEAPRSPPLA 1074  
 Db 1015 LVDAEEYLVPOQGFPCDPAAGAGMWHHRSSSTRSGGDLTLGLPSEEEAPRSPPLA 1074  
 Qy 1075 PSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOKYSDDPTVPLPSETDGVAPLTCSPQ 1134  
 Db 1075 PSEGAGSDVFDGDLGMAAKGLQSLPTHDPSPLOKYSDDPTVPLPSETDGVAPLTCSPQ 1136  
 Qy 1135 PEYVNOQDVPRPQPSRREGPLPAARPAATLIERAKTILSPKNGVYKXVFAFGAVENPEY 1194

Db 1137 PEYVNPQDVSRQPSREPGELPAPRPAGATLERPKTSLPGNGVVKVQVPAFGAVENPEY 1196  
 QY 1195 LTPGGGAAPQHPHPAPSPAPFNDLYWDDPPREGAAPSTFKGTPTAENPEYLGIDLPV 1253  
 Db 1197 LTPGGGAAPQHPHPAPSPAPFNDLYWDDPPREGAAPSTFKGTPTAENPEYLGIDLPV 1255

RESULT 12  
 AAE20479  
 ID AAE20479 standard; Protein, 1255 AA.  
 AC AAE20479;  
 XX 01-JUL-2002 (first entry)  
 DT 01-JUL-2002 (first entry)  
 XX Human Her-2/neu protein.  
 DE Human Her-2/neu protein; immune response; gene therapy; breast cancer;  
 KM human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.  
 XX Homo sapiens.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Region 1021..1030  
 FT /note="Naturally processed HLA-B44-restricted epitope"  
 XX MO200214503-A2.  
 XX 21-FEB-2002.  
 PF 14-AUG-2001; 2001MO-US41733.  
 PR 14-AUG-2000; 2000US-225152P.  
 PR 28-SEP-2000; 2000US-236428P.  
 PR 21-FEB-2001; 2001US-270520P.  
 XX (CORI-) CORIXA CORP.  
 PA Hand-Zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;  
 PI McNeill PD, Vedvick TS;  
 PI WPI: 2002-280758/32.  
 DR N-PSDB; AAD32743.  
 DR  
 XX Novel isolated Her-2/Neu polypeptide composition useful for therapy,  
 PT prevention and diagnosis of cancer, preferably breast cancer  
 XX  
 XX Disclosure; Page 114-117; 129p; English.  
 XX The invention relates to an isolated Her-2/Neu polypeptide composition  
 CC effective for eliciting an immune response. The invention is useful for  
 CC eliciting an immune response in a patient, where the patient is human  
 CC leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.  
 CC The composition is useful for the therapy and diagnosis of cancer.  
 CC Preferably breast cancer, in pharmaceutical compositions, e.g., vaccine  
 CC and other compositions for the diagnosis, prevention and treatment of  
 CC human malignancies, for stimulating and/or expanding T cells specific for  
 CC Her-2/Neu polypeptide and for inhibiting the development of cancer in a  
 CC patient. The invention is useful for stimulating a T cell response in a  
 CC human patient, as probe or primer for nucleic acid hybridization, to  
 CC selectively form duplex molecules with complementary stretches of the  
 CC entire Her-2/Neu gene or gene fragments of interest, to isolate a full  
 CC length gene from a suitable library, and to direct expression of a  
 CC polypeptide in appropriate host cells. The composition is useful in  
 CC prophylactic or therapeutic applications and for the treatment of cancer,  
 CC preferably for the immunotherapy of breast cancer and other Her-2/Neu-  
 CC associated malignancies. The invention is useful in gene therapy. The  
 CC present sequence is human Her-2/neu protein.  
 XX  
 SQ Sequence 1255 AA;

Best Local Similarity 96.7%; Pred. No. 0;  
 Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;  
 QY 1 MELAALCRGGLLALLPRGAATQVCTGTDMKLRPLASERTLDMRLHYOGGQVVOGSL 60  
 Db 1 MELAALCRGGLLALLPRGAATQVCTGTDMKLRPLASERTLDMRLHYOGGQVVOGSL 60  
 QY 61 ELTYLPTNASLSPLODIOGVGVYLIANQVAVPQRLIRVRGTQLPEDNVALVLDNG 120  
 Db 61 ELTYLPTNASLSPLODIOGVGVYLIANQVAVPQRLIRVRGTQLPEDNVALVLDNG 120  
 QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRNQLCYODTILMKDIFKXNOLA 180  
 Db 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQRNQLCYODTILMKDIFKXNOLA 180  
 QY 181 LTLIDNRSRACHPSPCKSGRCWSESEDDQSLTRTVACAGACAKPLPDDCCEOC 240  
 Db 181 LTLIDNRSRACHPSPCKSGRCWSESEDDQSLTRTVACAGACAKPLPDDCCEOC 240  
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMNPREGRTFGASCYTACP 300  
 Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMNPREGRTFGASCYTACP 300  
 QY 301 YNYLSTDVSGCTLVCEPLHNQEVTAEDGTORCEKSPCARVCGYGLCMQYIKANSKTIGT 360  
 Db 301 YNYLSTDVSGCTLVCEPLHNQEVTAEDGTORCEKSPCARVCGYGLCMQYIKANSKTIGT 360  
 QY 361 ELFPAGCKKIFGSLAFLESPFDGDPASNTAPLOPELOVFEETLEETGLYISAMPDLP 420  
 Db 361 ELFPAGCKKIFGSLAFLESPFDGDPASNTAPLOPELOVFEETLEETGLYISAMPDLP 420  
 QY 421 DLSVFONLOVIRGRILHNGAYSLLTQGLISWLGRLSLRELSGLALIHNNHLCFVHTV 480  
 Db 421 DLSVFONLOVIRGRILHNGAYSLLTQGLISWLGRLSLRELSGLALIHNNHLCFVHTV 480  
 QY 481 PMDQLEFRNPHOALLHNPANPEDECVGEGIAHQOLARCGMKPPGTQCNCSQFLRGQGC 540  
 Db 481 PMDQLEFRNPHOALLHNPANPEDECVGEGIAHQOLARCGMKPPGTQCNCSQFLRGQGC 540  
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCVACAHAHXPFFCVAR 600  
 Db 541 VEECRVLOGLPREYVNAHCLPCHPECOPONGSVTCFGEADQCVACAHAHXPFFCVAR 600  
 QY 601 PEGVAPDLSYMPIMKFPDEEGACOPCPINCHSCYVDLDDKCPADQASPLTSYSAVVG 660  
 Db 601 PEGVAPDLSYMPIMKFPDEEGACOPCPINCHSCYVDLDDKCPADQASPLTSYSAVVG 660  
 QY 661 ILVVVLGVVFGILIL-----FNNFVSPFWLVPKVSASHLEPLTPSGAMPQAOGRILK 714  
 Db 661 ILVVVLGVVFGILILKRQOKIRKTYMRLLOETEL---VEPLTPSGAMPQAOGRILK 714  
 QY 715 ETELKRVKVLGSAFGTYKGIWPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVVG 774  
 Db 715 ETELKRVKVLGSAFGTYKGIWPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVVG 774  
 QY 717 ETELKRVKVLGSAFGTYKGIWPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVVG 776  
 Db 717 ETELKRVKVLGSAFGTYKGIWPDGENVKIPVAIKVLRNTSPKANKEIIDEAYVVG 776  
 QY 775 VGSPPYSLRLGICLSTVQVLTQLMPYGLLDHVENGRISQDOLLMMQOIAAGMSVL 834  
 Db 775 VGSPPYSLRLGICLSTVQVLTQLMPYGLLDHVENGRISQDOLLMMQOIAAGMSVL 834  
 QY 835 EDVRLVHRDLAARVLYVSPNHVKITDGLARLIDITETVADGKVPKXMALESILR 894  
 Db 835 EDVRLVHRDLAARVLYVSPNHVKITDGLARLIDITETVADGKVPKXMALESILR 894  
 QY 895 REFTHOSDVWSGVWVWELMFGAKPYGIPAREIPDLLEKERRLPQPPICITIDVYMTIV 954  
 Db 895 REFTHOSDVWSGVWVWELMFGAKPYGIPAREIPDLLEKERRLPQPPICITIDVYMTIV 954  
 QY 955 KCMWIDSECRPRFRELIVSEFRMARDPORFVITQNEDELGPASPLDSTFYRSILJEDDMDG 1014  
 Db 955 KCMWIDSECRPRFRELIVSEFRMARDPORFVITQNEDELGPASPLDSTFYRSILJEDDMDG 1014  
 QY 1015 LVDABEYLVPOGFFCQPPDPAAGAGVHHRSSSTRGCGDILTGLPSEEBEAPRSLA 1074  
 Db 1015 LVDABEYLVPOGFFCQPPDPAAGAGVHHRSSSTRGCGDILTGLPSEEBEAPRSLA 1074

Query Match 96.9%; Score 6604; DB 23; Length 1255;



QY 775 VGSFVSRLLGICTSTVQVLTQMPYGCCLDHYENRGRGLSGDILMCMQIAKMSYL 834  
 Db 777 VGSFVSRLLGICTSTVQVLTQMPYGCCLDHYENRGRGLSGDILMCMQIAKMSYL 836  
 QY 835 EDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHAGGVPIKMALESILR 894  
 Db 837 EDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHAGGVPIKMALESILR 896  
 QY 895 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGEPLPQPICTIDVYIMV 954  
 Db 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGEPLPQPICTIDVYIMV 956  
 QY 955 KCMWIDSECRPRFRELVSFSSRMARDPQRFVYIIONEDLGASPLDSTFFRSLLDDMDG 1014  
 Db 957 KCMWIDSECRPRFRELVSFSSRMARDPQRFVYIIONEDLGASPLDSTFFRSLLDDMDG 1016  
 QY 1015 LVDAEYLVPQGGFFCDDPAPAGAGMHHRRSSSTSGGGDLTLGLEPSEBEAPRSPLA 1074  
 Db 1017 LVDAEYLVPQGGFFCDDPAPAGAGMHHRRSSSTSGGGDLTLGLEPSEBEAPRSPLA 1076  
 QY 1075 PSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOKYSBDPTVLPSETDGYVAPLTCSPQ 1134  
 Db 1077 PSEGAGSDVFDGDLGMAKGLQSLPTHDPSPLOKYSBDPTVLPSETDGYVAPLTCSPQ 1136  
 QY 1135 PEYVNOPVRRPSPREGLPAPAPAGATLERAKTSPGKGVYKVPFAFGAVENPEY 1194  
 Db 1137 PEYVNOPVRRPSPREGLPAPAPAGATLERAKTSPGKGVYKVPFAFGAVENPEY 1196  
 QY 1195 LTPOGGAAPQHPAPAPAFDNLVYWDOPPERGAPSTFKGTPTAENPEYGLDVPV 1253  
 Db 1197 LTPOGGAAPQHPAPAPAFDNLVYWDOPPERGAPSTFKGTPTAENPEYGLDVPV 1255

RESULT 14  
 ID A077114 standard; Protein: 1255 AA.  
 AC A077114:  
 DT 05-JUN-2002 (first entry)  
 XX Human Her-2/neu polypeptide.  
 DE Human Her-2/neu polypeptide.  
 XX Human; Her-2/neu; cytostatic; haematological malignancy; CML;  
 KM acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;  
 KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;  
 XX Hodgkin's lymphoma; T cell therapy.  
 OS Homo sapiens.  
 XX WO200213847-A2.  
 PN 21-FEB-2002.  
 PD 13-AUG-2001; 2001WO-US25408.  
 PF 14-AUG-2000; 2000US-0638280.  
 PR 28-SEP-2000; 2000US-0675904.  
 XX (COI-) CORIXA CORP.  
 PA Gaiger A, Cheever MA, Hand-zimmermann S;  
 PI WPI; 2002-280741/32.  
 DR N-PSDB; ABX10730.  
 XX Inhibiting haematological malignancy development by administering  
 PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide  
 PT encoding the polypeptide, or antigen presenting cells expressing the  
 PS polypeptide  
 XX Disclosure; Page 71-74; 74pp; English.

CC The invention relates to a method for inhibiting development of  
 CC haematological malignancy in a patient by administering a polypeptide  
 CC comprising an immunogenic portion of Her-2/neu or a polynucleotide  
 CC encoding the polypeptide. Antigen presenting cells that express the  
 CC protein can also be administered. The sequences are used for inhibiting  
 CC development of haematological malignancy such as acute myelogenous  
 CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic  
 CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's  
 CC lymphoma. This sequence represents the human Her-2/neu polypeptide.  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 96.9%; Score 6604; DB 23; Length 1255;  
 Best Local Similarity 96.7%; Pred. No. 0;  
 Matches 1218; Conservative 11; Mismatches 20; Indels 10; Gaps 2;  
 QY 1 METALCRGGLLALLPFGAATVCTGDMKRLPASETHLDMRLHYGGGVVQGNL 60  
 Db 1 METALCRGGLLALLPFGAATVCTGDMKRLPASETHLDMRLHYGGGVVQGNL 60  
 QY 61 ELTYLPTNLSLFLQDIQEVQGYVLAHQVQVPQRLRIYRGTOLEFDNYALAVLDNG 120  
 Db 61 ELTYLPTNLSLFLQDIQEVQGYVLAHQVQVPQRLRIYRGTOLEFDNYALAVLDNG 120  
 QY 121 DPLNNTPTVYTGASPGGLRELQRLSLTEILKGVLIORNPQLCYDITLWKDIFHKNOQLA 180  
 Db 121 DPLNNTPTVYTGASPGGLRELQRLSLTEILKGVLIORNPQLCYDITLWKDIFHKNOQLA 180  
 QY 181 LTLIDNRRACHPCGPMCKGRKMGESSEDDQSLTRYCAGGACRCKPLPTDCCHEC 240  
 Db 181 LTLIDNRRACHPCGPMCKGRKMGESSEDDQSLTRYCAGGACRCKPLPTDCCHEC 240  
 QY 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDPSMNPBEGRYTFGASCTYACP 300  
 Db 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDPSMNPBEGRYTFGASCTYACP 300  
 QY 301 YNYLSTDVSGCTLVGCILHNOEYTAEDGTGRCCKSPKRCVYCYGLMOYIKANSKITGIT 360  
 Db 301 YNYLSTDVSGCTLVGCILHNOEYTAEDGTGRCCKSPKRCVYCYGLMOYIKANSKITGIT 360  
 QY 361 ELTFAGCKKIFGSLAFLPESFDGDPASNTAPQPELOVFEFLTEITGLYISAMPDILP 420  
 Db 361 ELTFAGCKKIFGSLAFLPESFDGDPASNTAPQPELOVFEFLTEITGLYISAMPDILP 420  
 QY 421 DLSVFONLQVIRGILHNGAYSLTIOGUSTSWGLRSLRELGSGALLHHNTHLCFVHTV 480  
 Db 421 DLSVFONLQVIRGILHNGAYSLTIOGUSTSWGLRSLRELGSGALLHHNTHLCFVHTV 480  
 QY 481 PMDOLFRNPHOALLHTANRPEDECVGEGALCHQLCARHCWGPPTQCVNCSQPLRGQEC 540  
 Db 481 PMDOLFRNPHOALLHTANRPEDECVGEGALCHQLCARHCWGPPTQCVNCSQPLRGQEC 540  
 QY 541 VEECRVLOGLPREYVNAHRCLECHPECQONSGSYTCFGEADQCVACAHYKDPFVCAR 600  
 Db 541 VEECRVLOGLPREYVNAHRCLECHPECQONSGSYTCFGEADQCVACAHYKDPFVCAR 600  
 QY 601 PSGVYKPDLSYMPIMKPEDEGACQPCPINCNTSCTVDLDKGPAPQASPLTSISAVVG 660  
 Db 601 PSGVYKPDLSYMPIMKPEDEGACQPCPINCNTSCTVDLDKGPAPQASPLTSISAVVG 660  
 QY 661 ILVVYVLGVVFGILIL-----FNFVFSFWLREVPKVASHLEPLTPSGAMPQAOIRILK 714  
 Db 661 ILVVYVLGVVFGILILKRRQOKIRKYTMRLRLQETEL-----VEPLTPSGAMPQAOIRILK 716  
 QY 715 ETELKRVKVLGSGAGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKETIDEXYVWAG 774  
 Db 715 ETELKRVKVLGSGAGTYKGIWIPDGENVKIPVAIKYLRNTSPKANKETIDEXYVWAG 776  
 QY 775 VGSFVSRLLGICTSTVQVLTQMPYGCCLDHYENRGRGLSGDILMCMQIAKMSYL 834  
 Db 775 VGSFVSRLLGICTSTVQVLTQMPYGCCLDHYENRGRGLSGDILMCMQIAKMSYL 836  
 QY 835 EDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHAGGVPIKMALESILR 894

```

Db      ||| 837 EDVRLVHRDLAARNVLVSKPNHVKITDGLARLLDIDETEHADGGKVPIMWMALESILR 896
Qy      ||| 895 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMV 954
Db      ||| 897 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMV 956
Qy      ||| 955 KCMWIDSECRPRFRELVSFSRMAADPQRFVIVIONEDLGASPLDSTFYRSLLBEDDMDG 1014
Db      ||| 957 KCMWIDSECRPRFRELVSFSRMAADPQRFVIVIONEDLGASPLDSTFYRSLLBEDDMDG 1016
Qy      ||| 1015 LVDAEEYLVPOQGFPCDPAPAGAGMHHRRSSSTRSGGDLTLGLPSESEAPRSPLA 1074
Db      ||| 1017 LVDAEEYLVPOQGFPCDPAPAGAGMHHRRSSSTRSGGDLTLGLPSESEAPRSPLA 1076
Qy      ||| 1075 PSEGAAGDVFDGDLGMAAKGIQSLPTHDPSPLQRYSEDPTVPLPSETDGYVADLTCSPO 1134
Db      ||| 1077 PSEGAAGDVFDGDLGMAAKGIQSLPTHDPSPLQRYSEDPTVPLPSETDGYVADLTCSPO 1136
Qy      ||| 1135 PEYVNOPVRPOPSPRREGPLPAAPAGATLEBAKTUSPGKXGVKQDFARAGAVENPEY 1194
Db      ||| 1137 PEYVNOPVRPOPSPRREGPLPAAPAGATLEBAKTUSPGKXGVKQDFARAGAVENPEY 1196
Qy      ||| 1195 LTPQGAAPQHPHPAPSPADNLYWMDQDPERGAPSTFKGTPTAENPEYLGADV 1253
Db      ||| 1197 LTPQGAAPQHPHPAPSPADNLYWMDQDPERGAPSTFKGTPTAENPEYLGADV 1255

```

## RESULT 15

AAR39568 AAR39568 standard; Protein; 1433 AA.

AC AAR39568;

DT 07-FEB-1994 (first entry)

DE Sequence of c-erbB-2 tumour antigen.

XX Tumour antigen; c-erbB-2; glycoprotein.

XX Homo sapiens.

XX W09316185-A.

XX 19-AUG-1993.

XX 05-FEB-1993; 93WO-US01055.

XX 06-FEB-1992; 92US-0831967.

XX (CETU) CETUS ONCOLOGY CORP.

XX (CREA-) CREATIVE BIOMOLECULES INC.

XX Houston IL, Huston JS, Oppermann H, Ring DB;

XX MPI: 1993-272889/34.

XX N-PSDB: AAQ46083.

XX New single chain Fv polypeptide binding to C-erbB-2 tumour

XX antigen - for imaging or treating breast or ovarian cancer etc.

XX Disclosure; pages 48-54; 87dp; English.

XX c-erbB-2 refers to a protein antigen expressed on the surface of

XX tumour cells, such as breast and ovarian tumour cells, which is an

XX approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric

XX pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39568 represents

XX the location of a stop codon in AAQ46083.

XX Sequence 1433 AA;

Query Match 96.3%; Score 6561; DB 14; Length 1433;  
Best Local Similarity 96.2%; Pred. No. 0;

```

Matches 1211; Conservative 13; Mismatches 25; Indels 10; Gaps 2;
Qy      ||| 1 MEALACRWGLLALLPQGAATQVCTGDMKLRIPASETHLDMLRHLYQCGQVYQGNL 60
Db      ||| 1 MELAALCRWGLLALLPQGAATQVCTGDMKLRIPASETHLDMLRHLYQVQVQGNL 60
Qy      ||| 61 ELTYLPNLSLSTLQDIOEVQGVYLLAHNOVQVPLQRLRIYRGTLQFEDNYALAVDNG 120
Db      ||| 61 ELTYLPNLSLSTLQDIOEVQGVYLLAHNOVQVPLQRLRIYRGTLQFEDNYALAVDNG 120
Qy      ||| 121 DPLNNTTPYTGASPGGLRELOLSLTELILKGVLTQRNPOLCYQDTILWKDIFHNKNOA 180
Db      ||| 121 DPLNNTTPYTGASPGGLRELOLSLTELILKGVLTQRNPOLCYQDTILWKDIFHNKNOA 180
Qy      ||| 181 LTLIDNRSBACHPCSPCKGSGKSESSDDOSLTRYACAGGACAKPLPDCCHEOC 240
Db      ||| 181 LTLIDNRSBACHPCSPCKGSGKSESSDDOSLTRYACAGGACAKPLPDCCHEOC 240
Qy      ||| 241 AAGCTGPKSDCLACLFNHSIGCELAHCPALVTYNTDFESMNPREGRYTFGASCTYAC 300
Db      ||| 241 AAGCTGPKSDCLACLFNHSIGCELAHCPALVTYNTDFESMNPREGRYTFGASCTYAC 300
Qy      ||| 301 YNYLSTDVSGCTLVCEFLHNOEYTAEDGTORCEKSPCARVCYGLCMQYIKANSKRTGIT 360
Db      ||| 301 YNYLSTDVSGCTLVCEFLHNOEYTAEDGTORCEKSPCARVCYGLCMQYIKANSKRTGIT 360
Qy      ||| 361 ELFFAGCKKIFGSLAFLPSFGDDPASNTAPLOPEOLOVPELLEE: TGYLYISAMPDSL 420
Db      ||| 361 IOEPFAGCKRIFGSLAFLPSFGDDPASNTAPLOPEOLOVPELLEE: TGYLYISAMPDSL 420
Qy      ||| 421 DLSVFOQLQYIRGRILHNGAYSLTLQGLISWLGSLRELSGLALIHNNTHLSCFVHTV 480
Db      ||| 421 DLSVFOQLQYIRGRILHNGAYSLTLQGLISWLGSLRELSGLALIHNNTHLSCFVHTV 480
Qy      ||| 481 PMQDLFRNPHQALLHTANRDEDECEGGLAQGLCARGCMWGSPQYCNVCSQFLGQEC 540
Db      ||| 481 PMQDLFRNPHQALLHTANRDEDECEGGLAQGLCARGCMWGSPQYCNVCSQFLGQEC 540
Qy      ||| 541 VEECRVLOGLPREYVABHCLPCHPCCOPONGSVTCFGEADQCVACAHYKDPFCVAC 600
Db      ||| 541 VEECRVLOGLPREYVABHCLPCHPCCOPONGSVTCFGEADQCVACAHYKDPFCVAC 600
Qy      ||| 601 PSGVKPDLSTMPYWKPPDEBGAQCPQINCTHSCVDLDKGCABQASPLTSIVSAVVG 660
Db      ||| 601 PSGVKPDLSTMPYWKPPDEBGAQCPQINCTHSCVDLDKGCABQASPLTSIVSAVVG 660
Qy      ||| 661 ILVYVVLGVVFGILIKRQCKIRKYMRELQETEL-----VEPLTSGAMPNOQMRILK 714
Db      ||| 661 ILVYVVLGVVFGILIKRQCKIRKYMRELQETEL-----VEPLTSGAMPNOQMRILK 714
Qy      ||| 715 ETELARKVYLGSGAFGTVYKGIWIPGENYKIPVAIKVIRENTSPYANKEIDDEAYVMAG 774
Db      ||| 715 ETELARKVYLGSGAFGTVYKGIWIPGENYKIPVAIKVIRENTSPYANKEIDDEAYVMAG 774
Qy      ||| 775 VGSPPYRSLIGLICLTSTVOLVTQIMPYGCLDHYRENRGRLSGQDILLNMCQIAKMSYL 834
Db      ||| 775 VGSPPYRSLIGLICLTSTVOLVTQIMPYGCLDHYRENRGRLSGQDILLNMCQIAKMSYL 834
Qy      ||| 835 EDVRLVHRDLAARNVLVSKPNHVKITDGLARLLDIDETEHADGGKVPIMWMALESILR 894
Db      ||| 835 EDVRLVHRDLAARNVLVSKPNHVKITDGLARLLDIDETEHADGGKVPIMWMALESILR 894
Qy      ||| 895 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMV 954
Db      ||| 895 RRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYIMV 954
Qy      ||| 955 KCMWIDSECRPRFRELVSFSRMAADPQRFVIVIONEDLGASPLDSTFYRSLLBEDDMDG 1014
Db      ||| 955 KCMWIDSECRPRFRELVSFSRMAADPQRFVIVIONEDLGASPLDSTFYRSLLBEDDMDG 1014
Qy      ||| 1015 LVDAEEYLVPOQGFPCDPAPAGAGMHHRRSSSTRSGGDLTLGLPSESEAPRSPLA 1074
Db      ||| 1015 LVDAEEYLVPOQGFPCDPAPAGAGMHHRRSSSTRSGGDLTLGLPSESEAPRSPLA 1074

```

Qy 1075 PSEGAGSDVFPDGLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQ 1134  
Db 1077 PSEGAGSDVFPDGLGMAKGLQSLPTHDPSPLORYSEDPVPLPSETDGYVAPLTCSPQ 1136  
Qy 1135 PEYVNPDPVRPOPSPREGPLPAPAPAGATLERAKTLSPGKNGVXKDVFAFGAVENPEY 1194  
Db 1137 PEYVNPDPVRPOPSPREGPLPAPAPAGATLERAKTLSPGKNGVXKDVFAFGAVENPEY 1196  
Qy 1195 LTPQGGAAPQPHPPAFSPAFDNLXYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1253  
Db 1197 LTPQGGAAPQPHPPAFSPAFDNLXYWDQDPPERGAPSTFKGTPTAENPEYLGLDVPV 1255

Search completed: July 22, 2003, 09:16:38  
Job time : 41.875 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.3921 Seconds

(without alignments)  
5413.772 Million cell updates/sec

Title: SEQ4-695-709-14

Perfect score: 6853

Sequence: 1 MELAALCRWGLLALPDA.....TEKGTPTAENPEYLADVPV 1261

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database: 1: PIR:73:\*

2: PIR:\*

3: PIR:\*

4: PIR:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6641	96.9	1255	1 A24571	protein-tyrosine k
2	5844	85.3	1260	1 TVRINU	protein-tyrosine k
3	5834.5	85.1	1254	2 148161	P-185 precursor
4	3065	44.7	1210	1 GQHDE	epidermal growth f
5	3036	44.3	1210	2 A53183	epidermal growth f
6	3013.5	44.0	1223	1 TVCHLV	epidermal growth f
7	2884.5	42.1	1308	2 A47253	epidermal growth f
8	2592	37.8	1166	1 S06142	protein-tyrosine k
9	2334.5	34.1	1342	2 A36223	kinase-related tra
10	2248.5	32.8	1339	2 UC4387	epidermal growth f
11	1685.5	24.6	698	1 TVFVLV	protein-tyrosine k
12	1622	23.7	604	1 TVYUHL	protein-tyrosine k
13	1569	22.9	544	2 S35745	protein-tyrosine k
14	1562	22.8	545	2 S00727	kinase-related tra
15	1557.5	22.7	1330	1 GQFPE	epidermal growth f
16	1545	22.5	540	2 B44776	protein-tyrosine k
17	1543	22.0	540	1 TVFVEB	protein-tyrosine k
18	1509	22.0	644	2 A36325	epidermal growth f
19	1241	18.1	1333	2 E88257	protein-tyrosine k
20	1241	18.1	1374	2 S70712	protein-tyrosine k
21	1153	16.8	1369	2 S70713	protein-tyrosine k
22	1129	16.5	1717	1 A45558	epidermal growth f
23	1126	16.4	527	2 A42032	epidermal growth f
24	975.5	14.2	843	2 A27131	epidermal growth f
25	806.5	11.8	346	2 S13807	protein-tyrosine k
26	754.5	11.0	311	2 S13808	protein-tyrosine k
27	700	10.2	1363	2 T43220	insulin-like growth
28	670	9.8	1383	2 A36080	insulin receptor p
29	669.5	9.8	1372	2 A34157	insulin receptor p

30	668	9.7	1382	1 INHR	insulin receptor p
31	655	9.6	1477	2 T18534	protein-tyrosine k
32	653	9.5	1607	2 T43212	insulin-like growth
33	650.5	9.5	1300	2 A36302	insulin receptor-x
34	638	9.3	1268	2 B36502	insulin receptor-x
35	596	8.7	1367	1 IGHUR1	insulin-like growth
36	590	8.6	1390	2 T30346	insulin receptor -
37	588.5	8.6	2148	1 A56081	insulin receptor -
38	582	8.5	1301	2 S57245	insulin receptor (
39	577	8.4	1371	2 A33837	insulin-like growth
40	574	8.4	1114	1 S05582	protein-tyrosine k
41	571	8.3	987	2 A54092	protein-tyrosine k
42	569.5	8.3	952	2 I50612	protein-tyrosine k
43	566.5	8.3	984	2 A39753	protein-tyrosine k
44	562.5	8.2	977	2 S49004	tyrosine kinase Mp
45	556.5	8.1	976	2 A36355	protein-tyrosine k

## ALIGNMENTS

## RESULT 1

A24571 protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N:Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming protein  
C:Species: Homo sapiens (man)  
C>Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C:Accession: A24571; A25491; A44188; B44188; I59509; I57622

R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, Nature 319, 230-234, 1986

A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal gro

A:Reference number: A24571; MUID:86118663; PMID:3003577

A:Accession: A24571

A:Molecule type: mRNA

A:Residues: 1-1255 <YAM>

A:Cross-references: GB:M11767; NID:G182163; PIDN:CAA27060.1; PID:G31198

R:Semba, K.; Kanata, N.; Toyoshima, K.; Yamamoto, T.  
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/ep

A:Reference number: A25491; MUID:86016729; PMID:2995967

A:Accession: A25491

A:Molecule type: DNA

A:Residues: 737-1031 <SEM>

A:Cross-references: GB:M11767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R:Gustussens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seebu

A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares ch

A:Reference number: A44188; MUID:86070181; PMID:2999974

A:Accession: A44188

A:Molecule type: DNA

A:Residues: 740-910 <COU>

A:Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A:Accession: B44188

A:Molecule type: mRNA

A:Residues: 1-517, 'RALU', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

A:Cross-references: GB:M11730; NID:G183986

R:King, C.R.; Kraus, M.H.; Aaronson, S.A.  
Science 229, 974-976, 1985

A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.  
A:Reference number: I59509; MUID:85272597; PMID:2992089

A:Accession: I59509

A:Molecule type: DNA

A:Residues: 832-909 <REX>

A:Cross-references: GB:I29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R:Pal, K.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.  
Mol. Cell. Biol. 7, 2597-2601, 1987

A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcript

A:Reference number: I57622; MUID:87286898; PMID:3039351

A:Accession: I57622

A:Molecule type: DNA

A:Residues: 1-191 <TAU>

A:Cross-references: GB:M16792; NID:g183989; PID:AAA5637.1; PID:g553332  
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30  
 C:Genetics:  
 A:Gene: GDB:ERBB2; NGL; NEU; HER-2  
 A:Cross-references: GDB:120613; OMIM:164870  
 A:Map position: 17q21.1-17q21.1  
 A:Introns: 25/1; 75/3; 147/1; 883/3  
 A>Note: the list of introns is incomplete  
 C:Function:  
 A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 nase

F.1-31/Domain: signal sequence #status predicted <SIG>  
 F.122-1255/Product: protein-tyrosine kinase erbB2 #status predicted <MAT>  
 F.122-653/Domain: extracellular #status predicted <EXT>  
 F.70-304/Domain: EGF receptor extracellular domain repeat <EB1>  
 F.395-605/Domain: EGF receptor extracellular domain repeat <EB2>  
 F.654-675/Domain: transmembrane #status predicted <TM>  
 F.676-1255/Domain: intracellular #status predicted <INT>  
 F.718-983/Domain: protein kinase homology <KIN>  
 F.726-734/Region: protein kinase ATP-binding motif  
 F.68,124,187,253,530,571,629/Binding site: carboxylate (Asn) (covalent) #status predict  
 F.686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F.753/Active site: Lys #status predicted  
 F.1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 96.9% Score 6641; DB 1; Length 1255;  
 Best Local Similarity 97.3% Pred No.5.5e-266;  
 Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;

QY 1 MELALCRWGLLALPPGAASVCTGTDMKRLRLPASPEHLDMLRLHLYGGCCVVGNTL 60  
 DB 1 MELALCRWGLLALPPGAASVCTGTDMKRLRLPASPEHLDMLRLHLYGGCCVVGNTL 60  
 QY 61 ELTYPTNASTSLFODIOEGVYVLAHNOYRQVLRRLRVRGTCFEDNYALAVLDNG 120  
 DB 61 ELTYPTNASTSLFODIOEGVYVLAHNOYRQVLRRLRVRGTCFEDNYALAVLDNG 120  
 QY 121 DPLNNTTFTVGNAPGGLRELQRLSLTEILKGVLIQRPOLCYOQDTILMKDIFPKNNOLA 180  
 DB 121 DPLNNTTFTVGNAPGGLRELQRLSLTEILKGVLIQRPOLCYOQDTILMKDIFPKNNOLA 180  
 QY 181 LTLIDTNSRACHPCSPMKSCRCWGESSEDDQSLTRVYAGGACRCRGLPTCCHEQC 240  
 DB 181 LTLIDTNSRACHPCSPMKSCRCWGESSEDDQSLTRVYAGGACRCRGLPTCCHEQC 240  
 QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFEEMPNRGRYTFGASCVTACP 300  
 DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTFEEMPNRGRYTFGASCVTACP 300  
 QY 301 YNYLSTDVSGCTLVPLHNOEYTABDGTQRCCKSKPCARVYCYGLMOYIRANSKFTGIT 360  
 DB 301 YNYLSTDVSGCTLVPLHNOEYTABDGTQRCCKSKPCARVYCYGLMOYIRANSKFTGIT 360  
 QY 361 YNYLSTDVSGCTLVPLHNOEYTABDGTQRCCKSKPCARVYCYGLMOYIRANSKFTGIT 360  
 DB 361 YNYLSTDVSGCTLVPLHNOEYTABDGTQRCCKSKPCARVYCYGLMOYIRANSKFTGIT 360  
 QY 420 ELLEPGCKKIFGSLALFPESEFDPGASNTAPLQEPQLQVETLEETITGLYSAMPDLSL 420  
 DB 420 ELLEPGCKKIFGSLALFPESEFDPGASNTAPLQEPQLQVETLEETITGLYSAMPDLSL 420  
 QY 421 IOEFGCKKIFGSLALFPESEFDPGASNTAPLQEPQLQVETLEETITGLYSAMPDLSL 420  
 DB 421 IOEFGCKKIFGSLALFPESEFDPGASNTAPLQEPQLQVETLEETITGLYSAMPDLSL 420  
 QY 481 PMDQLEFRNPKHALLTANRPEDECGEGELACHOLCARHCHGPGSTOCVNSOFLRGDEC 540  
 DB 481 PMDQLEFRNPKHALLTANRPEDECGEGELACHOLCARHCHGPGSTOCVNSOFLRGDEC 540  
 QY 541 VEEGRVLOGLPREYVNAHCLPHEPCQPNQSGVTCFSPREADQCVACAHYKDPFCVAVRC 600  
 DB 541 VEEGRVLOGLPREYVNAHCLPHEPCQPNQSGVTCFSPREADQCVACAHYKDPFCVAVRC 600  
 QY 601 PSGVPRDLSYMPIMKFPDEEGACQPCPINCHTSCVDLDDKCGPAGQASPLTISVAVNG 660  
 DB 601 PSGVPRDLSYMPIMKFPDEEGACQPCPINCHTSCVDLDDKCGPAGQASPLTISVAVNG 660

DB 601 PSGVPRDLSYMPIMKFPDEEGACQPCPINCHTSCVDLDDKCGPAGQASPLTISVAVNG 660  
 QY 661 ILVVVLGVVFGILLIKRQOKIRKTYMRLLLOETELVEVLTSSGAMPNQAQRILKEFNK 720  
 DB 661 ILVVVLGVVFGILLIKRQOKIRKTYMRLLLOETELVEVLTSSGAMPNQAQRILKEFNK 720  
 QY 721 FTYSFWLRYKVSAS-HLETVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAVY 779  
 DB 721 FTYSFWLRYKVSAS-HLETVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAVY 779  
 QY 779 FTYSFWLRYKVSAS-HLETVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAVY 779  
 DB 779 FTYSFWLRYKVSAS-HLETVYKGIWIPGENVKIPVAIKVIRENTSPKANKEILDEAVY 779  
 QY 840 SYLEDVRLVHRDLAAENVLVKSPNHVKINDPGLARLDDIDEVYHADGGKVPKMMALS 899  
 DB 840 SYLEDVRLVHRDLAAENVLVKSPNHVKINDPGLARLDDIDEVYHADGGKVPKMMALS 899  
 QY 899 SYLEDVRLVHRDLAAENVLVKSPNHVKINDPGLARLDDIDEVYHADGGKVPKMMALS 899  
 DB 899 SYLEDVRLVHRDLAAENVLVKSPNHVKINDPGLARLDDIDEVYHADGGKVPKMMALS 899  
 QY 900 ILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPPQPICTIDVYM 959  
 DB 900 ILRRRTTHOSDWSYGVTVWELMTFGAKPYDGIPAREIPDLLEKGERLPPQPICTIDVYM 959  
 QY 959 IMYKCMWIDSECRPRRELVSERSRPAARDPQRFVYQNEDLGASPLDSTFYRSLLED 1013  
 DB 959 IMYKCMWIDSECRPRRELVSERSRPAARDPQRFVYQNEDLGASPLDSTFYRSLLED 1013  
 QY 1020 MGDLVAEEFLVPOQGFCDPAPAGAWVHRHRSSTRSGGDLTLGLEPSEEARPS 1079  
 DB 1020 MGDLVAEEFLVPOQGFCDPAPAGAWVHRHRSSTRSGGDLTLGLEPSEEARPS 1079  
 QY 1079 PLAPSEGAGSDVFDGGLMGCAKGLQSLTTHPSPLQRYSEDPVLPSETDGYVAPLTC 1133  
 DB 1079 PLAPSEGAGSDVFDGGLMGCAKGLQSLTTHPSPLQRYSEDPVLPSETDGYVAPLTC 1133  
 QY 1140 SPQEVYNQPDVAPQPPSPREGPLPAARPAAGATLEPKILSPKAGVMDVAFGAVEN 1193  
 DB 1140 SPQEVYNQPDVAPQPPSPREGPLPAARPAAGATLEPKILSPKAGVMDVAFGAVEN 1193  
 QY 1200 PEYLTPQGAAPQHPPEPAPFAFDNLVYWDQDPPRGAAPSTFKGTPTAENBEVGLDV 1259  
 DB 1200 PEYLTPQGAAPQHPPEPAPFAFDNLVYWDQDPPRGAAPSTFKGTPTAENBEVGLDV 1259  
 QY 1259 PEYLTPQGAAPQHPPEPAPFAFDNLVYWDQDPPRGAAPSTFKGTPTAENBEVGLDV 1259  
 DB 1259 PEYLTPQGAAPQHPPEPAPFAFDNLVYWDQDPPRGAAPSTFKGTPTAENBEVGLDV 1259  
 QY 1260 PV 1261  
 DB 1260 PV 1261  
 QY 1254 PV 1255  
 DB 1254 PV 1255

RESULT 2  
 TIRNTU  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 31-Dec-1988 #sequence, revision 31-Dec-1988 #text, change 11-Jun-1999  
 C:Accession: A24562; A61204  
 R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.  
 Nature 319, 226-230, 1986  
 A:Title: The neu oncogene encodes an epidermal growth factor receptor-related prote  
 A:Reference number: A24562; M01D:86118662; PMID:3345311  
 A:Accession: A24562  
 A:Molecule type: mRNA  
 A:Residues: 1-1260 <BAR>  
 A:Cross-references: EMBL:X03362; NID:g56745; PID:CAA27059.1; PID:g56746  
 R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; C  
 Carling, D. Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals  
 2-thiazolyl]formamide or N-methyl-N-nitrosourea.  
 A:Reference number: A61204; M01D:92035293; PMID:1682063  
 A:Accession: A61204  
 A:Molecule type: DNA  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 637-663, 'V', 665-702 <MAS>  
 A>Note: authors translated the codon GCA for residue 25 as Val  
 C:Genetics:  
 A:Gene: neu



C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 F:1-13/Domain: signal sequence #status predicted <Sig>  
 F:20-1260/Product: protein-tyrosine kinase neu #status predicted <Mat>  
 F:658-680/Domain: transmembrane #status predicted <TM>  
 F:723-988/Domain: protein kinase homology <Kin>  
 F:731-739/Region: protein kinase ATP-binding motif  
 F:721-191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:691/Binding site: phosphate (Thr) (covalent) #status predicted  
 F:758/Active site: lys #status predicted  
 F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 85.3%; Score 5844; DB 1; Length 1260;  
 Best Local Similarity 85.7%; Pred. No. 3,2e-233;  
 Matches 1083; Conservative 52; Mismatches 119; Indels 10; Gaps 4;

QY 1 MELALCRWGLLALLPPGAASVQVCTGTDKRLPASPETHLDMRLHYGCCVYQGNL 60  
 DB 4 MELAAWCRWGLLALLPPGIAGVCTGTDKRLPASPETHLDMRLHYGCCVYQGNL 63  
 QY 61 ELTYLPTNASLFLDIOEVQGYVLIANQVROVPLQRLIRVGTQLFEDNYALAVDNG 120  
 DB 64 ELTYLPANASLSTLODIOEVQGYMLIANQVRAVPLQRLIRVGTQLFEDKXALAVDNR 123  
 QY 121 DPLNNTTPTVT-GASPGGLREQLRS:TEILKGGVLIQRNPOLCYODTIIMKDIFFKNNQL 179  
 DB 124 DPODNVAASTPGRTPEGLRELQLRS:TEILKGGVLIQRNPOLCYODTMVIMKDVFRKNQL 183  
 QY 180 ALTLIDTNSRACHPSPCKGSRGSESEDDQS:LTRVCAAGGACRGKRP:PTCCHEQ 239  
 DB 184 APVDIDTNSRACPPAPACKDNHCWGESPEDCQILTGITCTSGCARCGRLPTDCHEQ 243  
 QY 240 CAAGCTGPRHSDCLACIAFNHSGICELHCPALVTYNTDFESMPNEGRYTFGASCVTAC 299  
 DB 244 CAAGCTGPRHSDCLACIAFNHSGICELHCPALVTYNTDFESMPNEGRYTFGASCVTTC 303  
 QY 300 PNYVISTDVGSCITVCEPLNQEYTAEDGRCCKSKPCARVCYGIQMGYIIRANKFQI 359  
 DB 304 PNYVISTEVSCTVCPNNQYTAEDGRCCKSKPCARVCYGIQMGYIIRANKFQI 363  
 QY 360 TELEFAGCKKIFGSLAFLPESPDSPASNTAPLOPEQLQVFELEETITGYLISAMPDL 419  
 DB 364 NYGEFPGCKKIFGSLAFLPESPDSPASNTAPLOPEQLQVFELEETITGYLISAMPDL 423  
 QY 420 PDLVSQNTQVIRGRILHNGAYSILTIQGISWLGIRSLRELSGIALIHNHTHLCFVHT 479  
 DB 424 RDLVSQNTQVIRGRILHNGAYSILTIQGISWLGIRSLRELSGIALIHNHTHLCFVHT 483  
 QY 480 VPMDQLFRNPQALLTANRPEDE-CVGEGLACHQLCARAGHCGMPGTQCVNCSQFLRG 538  
 DB 484 VPMDQLFRNPQALLHSGNRPEDELCVSSGLVYNSLCAIGHCMGPRGTQCVNCSHFLRG 543  
 QY 539 ECVEECRVLQGLPREVYNAHCLPCHEPCOPONGSVTCGPPEADQCVCAHYKDPFCVA 598  
 DB 544 ECVEECRVLQGLPREVYNAHCLPCHEPCOPONGSVTCGPPEADQCVCAHYKDPFCVA 603  
 QY 599 RCPSSGKPPDLSTYMPKPPDEBGAOCPCPINCCHSVDDDDGCPAPKASPLTSISAV 658  
 DB 604 RCPSSGKPPDLSTYMPKPPDEBGAOCPCPINCCHSVDDDDGCPAPKASPLTSISAV 663  
 QY 659 VGLILVAVLVGFGILIKRQOKIRKYTNRLLOETELVEPLTPSGAMPNOQAMRLKEF 718  
 DB 664 EGVLLFLILVAVLVGILIKRQOKIRKYTNRLLOETELVEPLTPSGAMPNOQAMRLKET 723  
 QY 719 NNFTVAFWLRVPVYAS-SLETTYKGIWT PDGENYKI PAIKVLRNTPSPKANKETLDA 777  
 DB 724 E-----LRKYVVGSGAFGYIKGIW PDGENYKI PAIKVLRNTPSPKANKETLDA 776  
 QY 778 YNAGVGSFVYSLIGICITSTVQVLTQMLPYGCLLDHVERNGRLSGDLLNMCWQIAK 837  
 DB 777 YNAGVGSFVYSLIGICITSTVQVLTQMLPYGCLLDHVERNGRLSGDLLNMCWQIAK 836  
 QY 838 GMSYLEDVALVHRDLAARNVLYKSPNHVITITDEGLARLLDIDETVHADGKVPYIKMML 897

DB 837 GMSYLEDVALVHRDLAARNVLYKSPNHVITITDEGLARLLDIDETVHADGKVPYIKMML 896  
 QY 898 ESILRRRFTHOSDWSVGVYTWELMTGAKPYDGI PARBEIPDLLEKGERLPQPICTIDV 957  
 DB 897 ESILRRRFTHOSDWSVGVYTWELMTGAKPYDGI PARBEIPDLLEKGERLPQPICTIDV 956  
 QY 958 YMIWKCMIIDSECRPPRELVSFSPKAPRPFVYVIONEDLPASPDLSTFRSLLED 1017  
 DB 957 YMIWKCMIIDSECRPPRELVSFSPKAPRPFVYVIONEDLPASPDLSTFRSLLED 1016  
 QY 1018 DDMGDVDAEEYLVPOGFFCPDPAAGQVHHRRHSSSTRSGGDTITLGLPSEEPAP 1077  
 DB 1017 DDMGDVDAEEYLVPOGFFCPDPAAGQVHHRRHSSSTRSGGDTITLGLPSEEPAP 1076  
 QY 1078 RSPAPSGASDVFDGDLGCAKGIQSLPTHPSPQLQXSEPTVPLSEDTGYVAPL 1137  
 DB 1077 RSPAPSGASDVFDGDLGCAKGIQSLPTHPSPQLQXSEPTVPLSEDTGYVAPL 1136  
 QY 1138 TCSPOPEYVNPDPVPPSPREGPLPAARPAGATLEAKTSLSPKNGVVDVAFGAV 1197  
 DB 1137 ACSPOPEYVNPDPVPPSPREGPLPAARPAGATLEAKTSLSPKNGVVDVAFGAV 1196  
 QY 1198 ENPEYLTPOGGAAPQHPPPAFDNLVYWDQDPPERCAPPESTFKTPTAENPEYGL 1257  
 DB 1197 ENPEYLTPOGGAAPQHPPPAFDNLVYWDQDPPERCAPPESTFKTPTAENPEYGL 1256  
 QY 1258 DVPV 1261  
 DB 1257 DVPV 1260

## RESULT 3

148161  
 P-185 precursor - golden hamster  
 C:Species: Mesocricetus auratus (golden hamster)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 18-Jun-1999  
 R:Nakamura, T.; Uehijima, T.; Ishikawa, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishi  
 Gene 140, 251-255, 1994  
 A>Title: Cloning and activation of the Syrian hamster neu proto-oncogene.  
 A:Reference number: 148161; MUID:94193007; PMID:7908275  
 A:Accession: 148161  
 A>Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1254 <RES>  
 A:Cross-references: GB:D16295; NID:G493236; PID:BAA03801.1; PID:G747595  
 C:Genetics:  
 A:Gene: neu  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP  
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.1%; Score 5834.5; DB 2; Length 1254;  
 Best Local Similarity 85.3%; Pred. No. 7,8e-233;  
 Matches 1077; Conservative 61; Mismatches 115; Indels 9; Gaps 3;

QY 1 MELALCRWGLLALLPPGAASVQVCTGTDKRLPASPETHLDMRLHYGCCVYQGNL 60  
 DB 1 MELAAWCRWGLLALLSPGASGTQVCTGTDKRLPASPETHLDMRLHYGCCVYQGNL 60  
 QY 61 ELTYLPTNASLFLDIOEVQGYVLIANQVROVPLQRLIRVGTQLFEDNYALAVDNG 120  
 DB 61 ELTYLPANASLSTLODIOEVQGYMLIANQVRAVPLQRLIRVGTQLFEDKXALAVDNR 120  
 QY 121 DPLNNTTPTVTGASPGGLREQLRS:TEILKGGVLIQRNPOLCYODTIIMKDIFFKNNQLA 180  
 DB 121 DPLNNTTPTVTGASPGGLREQLRS:TEILKGGVLIQRNPOLCYODTIIMKDVFRKNQLA 180  
 QY 181 LTLIDTNSRACHPSPCKGSRGSESEDDQS:LTRVCAAGGACRGKRP:PTCCHEQ 240  
 DB 181 LTLIDTNSRACPPAPACKDNHCWGESPEDCQILTGITAPRAVPAARLPTDCHEQ 240

QY 241 AAGCTGPKHSDDLACIHFHNSGICELHCPALVTYNTDTESMNPBGRTTFGASCTTAC 300  
 Db 241 AAGCTGPKHSDDLACIHFHNSGICELHCPALVTYNTDTESMNPBGRTTFGASCTTAC 300  
 QY 301 YNLTSTDVSGCTVCPHNOEYTAEDGTORCEKSPKACVCGLMQVYIKANSKFIG:IT 360  
 Db 301 YNLTSTDVSGCTVCPHNOEYTAEDGTORCEKSPKACVCGLMQVYIKANSKFIG:IT 360  
 QY 361 EDEPAGCKTIFGSLALPESFDGDPASNTAPLOPEOLQVETLEETITGUYISAMPDCLP 420  
 Db 361 EDEPAGCKTIFGSLALPESFDGDPASNTAPLOPEOLQVETLEETITGUYISAMPDCLP 420  
 QY 421 DLSPFQNLQVIRIRIILHNGAYSLTLOGLGSMGLSLRLSGLLHNNHNLGVHNV 480  
 Db 421 DLSPFQNLQVIRIRIILHNGAYSLTLOGLGSMGLSLRLSGLLHNNHNLGVHNV 480  
 QY 481 PMDQLEFNPHQALLHTANRDEDECVBEGGLAQHOLCARHGWGPGPTQVNCQOFLRGQEC 540  
 Db 481 PMDQLEFNPHQALLHTANRDEDECVBEGGLAQHOLCARHGWGPGPTQVNCQOFLRGQEC 540  
 QY 541 VEEGCRVLOGLPREYVNAHCLPHPEGQPNQSVTQEGPADQCVCAHYKDPFCVAC 600  
 Db 541 VEEGCRVLOGLPREYVNAHCLPHPEGQPNQSVTQEGPADQCVCAHYKDPFCVAC 600  
 QY 601 PSGVVPDLSTYMPIWKPPDEBGAQCPQPINCTHSVDLDDKCEBQASPLTSIVSAVVG 660  
 Db 601 PSGVVPDLSTYMPIWKPPDEBGAQCPQPINCTHSVDLDDKCEBQASPLTSIVSAVVG 660  
 QY 661 ILVLVGLVVFGLIKRQOKIKRYTMRLLQETELVEPLTSGAMPNQAQMLKEPNK 720  
 Db 661 ILVLVGLVVFGLIKRQOKIKRYTMRLLQETELVEPLTSGAMPNQAQMLKEPNK 720  
 QY 721 FTVSFWLVRVKSAS-HLETVYKGIWIPDGENVKIPVALKYLENTPSPANKELDEAVY 779  
 Db 721 FTVSFWLVRVKSAS-HLETVYKGIWIPDGENVKIPVALKYLENTPSPANKELDEAVY 779  
 QY 780 MAGVSPYVRLIGTICITSTVOLVQTMPIGCLLDHYRENRGLSGQDLINMCQIAKGM 839  
 Db 780 MAGVSPYVRLIGTICITSTVOLVQTMPIGCLLDHYRENRGLSGQDLINMCQIAKGM 839  
 QY 834 SYLEDVRLVHRDLAARNAVYKSPNHVKITDFGLARLLDIDETIYHADGKVPKIALES 893  
 Db 834 SYLEDVRLVHRDLAARNAVYKSPNHVKITDFGLARLLDIDETIYHADGKVPKIALES 893  
 QY 894 ILRRRFTHQSDVMSYGTWELMTFFGAKPYDGIPIAREBIPDLLEKGRRLPQPICTIDVYM 953  
 Db 894 ILRRRFTHQSDVMSYGTWELMTFFGAKPYDGIPIAREBIPDLLEKGRRLPQPICTIDVYM 953  
 QY 954 IMYKCMWIDSECRPRFELTSEFSRMAKDPQRFVJIONEDLGPASPLDSFYRSLLEDD 1019  
 Db 954 IMYKCMWIDSECRPRFELTSEFSRMAKDPQRFVJIONEDLGPASPLDSFYRSLLEDD 1019  
 QY 1020 MGDVLDAEYLVPOQGFCDPAPAGAGVWHRHRSSTSTRSGGDLTLGLPSEBAPRS 1079  
 Db 1020 MGDVLDAEYLVPOQGFCDPAPAGAGVWHRHRSSTSTRSGGDLTLGLPSEBAPRS 1079  
 QY 1079 MGDVLDAEYLVPOQGFCDPAPAGAGVWHRHRSSTSTRSGGDLTLGLPSEBAPRS 1079  
 Db 1079 MGDVLDAEYLVPOQGFCDPAPAGAGVWHRHRSSTSTRSGGDLTLGLPSEBAPRS 1079  
 QY 1080 PLAPBEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDETVLPSSTQVYAPLTC 1139  
 Db 1080 PLAPBEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDETVLPSSTQVYAPLTC 1139  
 QY 1139 PLAPBEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDETVLPSSTQVYAPLTC 1139  
 Db 1139 PLAPBEGAGSDVFDGDLGMAAGLQSLPHTDPSPLQRYSEDETVLPSSTQVYAPLTC 1139  
 QY 1140 SPQEVYNQDVPARPPSPREGLPAAPAGATLEBRAKTLSPKNGVVDVAFGAVEN 1199  
 Db 1140 SPQEVYNQDVPARPPSPREGLPAAPAGATLEBRAKTLSPKNGVVDVAFGAVEN 1199  
 QY 1199 SPQEVYNQDVPARPPSPREGLPAAPAGATLEBRAKTLSPKNGVVDVAFGAVEN 1199  
 Db 1199 SPQEVYNQDVPARPPSPREGLPAAPAGATLEBRAKTLSPKNGVVDVAFGAVEN 1199  
 QY 1200 PEYITPQGAAPQHPAPASPAFDNIYVWDQDPERGAPSPFKGTPTANENYGLDV 1259  
 Db 1200 PEYITPQGAAPQHPAPASPAFDNIYVWDQDPERGAPSPFKGTPTANENYGLDV 1259  
 QY 1259 PEYITPQGAAPQHPAPASPAFDNIYVWDQDPERGAPSPFKGTPTANENYGLDV 1259  
 Db 1259 PEYITPQGAAPQHPAPASPAFDNIYVWDQDPERGAPSPFKGTPTANENYGLDV 1259

## RESULT 4

GQHV  
 epidermal growth factor receptor precursor - human  
 N:Contains: protein-tyrosine kinase (BC 2.7.1.112) erbB  
 M:Species: Homo sapiens (man)  
 C:Date: 15-Nov-1994 #sequence revision 27-Nov-1995 #text change 11-Jun-1999  
 A:Accession: A00641; A25772; S30024; A36672; A00642; A43615; A23062; A05281; A60143  
 R:Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee,  
 Xg, P.H.  
 Nature 309, 418-425, 1984  
 A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expressi  
 A:Reference number: A00641; MUID:84219729; PMID:6326312  
 A:Accession: A00641  
 A:Molecule type: mRNA  
 A:Residues: 1-1210 <U>L>  
 A:Cross-references: EMBL:X00588; NID:G31113; PIDN:CAA25240.1; PID:G757924  
 A:Note: the authors translated the codon AAG for residue 540 as Asn  
 R:Rishi, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.  
 Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985  
 A:Title: Characterization and sequence of the promoter region of the human epiderma  
 A:Reference number: A25772; MUID:85270438; PMID:2991899  
 A:Accession: A25772  
 A:Status: translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-29 <ISH>  
 A:Cross-references: GB:M1124; NID:G181981; PIDN:AAA52370.1; PID:G553272  
 R:Halley, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ulrich, A.; Waterfield, M.  
 Oncogene Res. 1, 375-396, 1987  
 A:Title: The human EGF receptor gene: structure of the 110 kb locus and identifiat  
 A:Reference number: S30024; MUID:88217333; PMID:3329716  
 A:Accession: S30024  
 A:Molecule type: DNA  
 A:Residues: 1-29 <HA2>  
 A:Cross-references: EMBL:X06370; NID:G31118; PIDN:CAA2668.1; PID:G31119  
 R:Halley, J.D.; Waterfield, M.D.  
 J. Biol. Chem. 266, 1746-1753, 1991  
 A:Title: Contributory effects of de Novo transcription and premature transcript ter  
 A:Reference number: A36672; MUID:91107677; PMID:1988448  
 A:Accession: A36672  
 A:Molecule type: DNA  
 A:Residues: 1-29 <HAL>  
 A:Cross-references: GB:M8425; NID:G181977; PIDN:AAA63171.1; PID:G553271  
 R:Halley, J.D.; Waterfield, M.D.  
 J. Biol. Chem. 266, 1746-1753, 1991  
 A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of  
 A:Reference number: A00642; MUID:84245835; PMID:6330563  
 A:Accession: A00642  
 A:Molecule type: mRNA  
 A:Residues: RCMWRA, 150-187, 'KSVIQAV', 195, 'M', 197, 'A', 199-222, 'S', 224-304, 'RA', 30  
 ', 798-799, 'TD', 802-811, 'R', 813-942 <XOY>  
 A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF  
 R:Rishi, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.  
 Science 224, 843-848, 1984  
 A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplifica  
 A:Reference number: A43615; MUID:84196372; PMID:6326261  
 A:Accession: A43615  
 A:Molecule type: mRNA  
 A:Residues: 713-964 <LIN>  
 A:Experimental source: epidermoid carcinoma cell line A431  
 R:Stimmen, F.A.; Gope, M.L.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.I.  
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984  
 A:Reference number: A23062; MUID:85046483; PMID:6093780  
 A:Accession: A23062  
 A:Molecule type: mRNA  
 A:Residues: 1028-1210 <SIM>  
 R:Weber, M.; Gyll, G.N.; Speiss, J.  
 Science 224, 294-297, 1984  
 A:Reference number: A05281; MUID:84172183; PMID:6324343  
 A:Accession: A05281  
 A:Molecule type: protein  
 A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>

R:Russo, M.W.; Lukasz, T.J.; Cohen, S.; Stamos, J.V.  
 J. Biol. Chem. 260, 5205-5208, 1985  
 A:Title: Identification of residues in the nucleotide binding site of the epidermal growth factor receptor  
 A:Reference number: A60143; PMID:85182650; PMID:2985580  
 A:Accession: A60143  
 A:Molecule type: Protein  
 A:Residues: 740-744, 'X', 746-747 <RUS>  
 R:Moczekowski, B.; Mosig, G.; Cohen, S.  
 Nature 309, 270-273, 1984  
 A:Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide dismutase  
 A:Reference number: A38023; PMID:84191554; PMID:6325948  
 A:Contents: annotation; receptor activity  
 A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA  
 R:Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.  
 Cell 59, 33-43, 1989  
 A:Title: Functional independence of the epidermal growth factor receptor from a domain I  
 A:Reference number: A33331; PMID:90003233; PMID:2790960  
 A:Contents: annotation; internalization signal  
 C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor complex  
 C:Genetics:  
 A:Gene: GDB:EGFR  
 A:Cross-references: GDB:120610; OMIM:131550  
 A:Map position: 7p12.3-7p12.1  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 F:1,24/Domain: signal sequence #status predicted <SIG>  
 F:1,24/Domain: signal sequence #status predicted <SIG>  
 F:25-64/Domain: extracellular #status predicted <EXT>  
 F:25-64/Domain: extracellular #status predicted <EXT>  
 F:75-300/Domain: EGF receptor extracellular domain repeat <E2>  
 F:390-600/Domain: EGF receptor extracellular domain repeat <E2>  
 F:666-668/Domain: transmembrane #status predicted <TM>  
 F:669-1210/Domain: intracellular #status predicted <INT>  
 F:110-975/Domain: protein kinase homology <KIN>  
 F:118-726/Region: protein kinase ATP-binding motif  
 F:999-1046/Region: coated-pit mediated internalization signal  
 F:1047-1210/Region: inhibitory  
 F:128,175,352,413,444,528,603/Binding site: carbohydrate (Asn) (covalent) #status predicted  
 F:745/Active site: Lys #status experimental

Query Match 44.7%; Score 3065; DB 1; Length 1210;  
 Best Local Similarity 48.7%; Pred. No. 5,4e-119;  
 Matches 620; Conservative 179; Mismatches 359; Indels 114; Gaps 24;

```

QY 11 LLLALLPPGAA--STOVCTGTDWKLRLPASPTHTLMDLRLHLYGCGVCGVCGNLETTLPN 68
DB 14 LLLALCPASRALEEKVCCGTSNKLQGLGFEDHFLSLQMFNCCVVLGNLETTVQGN 73
QY 69 ASLSPLQDIOEVGVYLAHNOYRQVPLQRLIVRGSTQFEDNYALAVLDNGDPLNNTTP 128
DB 74 YDLSFLKTIQEVAGVYLAHNTVERIPLENTLQIRGNMYEENSVALAVLSND----- 126
QY 129 VFGASPGSLREILQSLTELKGVLIQGNPOLCYQDITLMMDIFHKNNQALTLIDTNR 188
DB 127 ---ANKTGLKELMERLQELHGAIVFSSNPALCVNESIQMRDLYVSDPLSNNSMDFQNH 183
QY 189 SRACHPQSPWCKSRGCGESSDDCSLITRVCAAGCA-RCKGPLPTDCCHQCAAGCTGP 247
DB 184 LGSQCKDCPSCPGSCWGAEBEWCCKLTIKLCAQCSGRCRGKSPDCCHQCAAGCTGP 243
QY 248 KNSDCLACLHFNHSGICELHCPALVYNDTFESMNPGRRTFGASCCTTACPYNLTSD 307
DB 244 RESDCLVCRKFRDEATCKTQCPPLMKNPTTYQMDVNPFGKYSFGATCVKCKPRNVVVD 303
QY 308 VGSCTLVCPILHNOEVAEDGTORCEKSPKCAVCGYLMQVYKANSKFIQTELE-FAG 366
DB 304 HGSCTVACGADSYEM-EDGVKCKCKCEGCRVYCGIGIERK-DSLSINATNTHFGN 361
QY 367 CKKIFGSLAFLEPSFPGDPASNTAPLQPELOVFEFTLEITGYLYISAMPDLPDLISVQ 426
DB 362 CTSISGDHLPLPAFGRGDSFTHTPPDPELDILKTKVKEITGILLQAMPENRTIDHAF 421
QY 427 NLOVIRGLIHNGAVSLTLOGLSIWLGLRSLEISGLALHNNHNLGFVHTVPDOLF 486

```

```

DB 422 NLEIRGRTKOHQGSFLAVSLNITSLRLSLKEISDDVILSGNKLQYANTIMKKLF 481
QY 487 RNPQALHTNRPREDCEVGGGLCHQICARHGCHGPGPTQCVNCSQFLPQGECEBGRV 546
DB 482 GTSQCKTKLISNRGENSKATGQVCHALCSBGCMPBPPOCVSQRNYSRGREVDCKL 541
QY 547 LOGLPREYVNAHCLPCHPECPONGSVTCEGPBDAQVACAHYKDPFCVACRSGVVP 606
DB 542 LEGEPREYVNESECIOCHPECLPQAMNITTCGRGPDNCTIQCAHYIDGHCVTCPAGWG 601
QY 607 DLSTYPIKFPDEBACQCPPLNCHSCVDDDDKCPBDAQASPLTISVAVG--ILL 663
DB 602 ENNTL-VKTYADAGVCHLCHPNCCTYGTGPGLECCPNGKIP--SIATGAGALLLL 658
QY 664 VVLDGVVFGILIKRQCKIRKYTRRLQETELVEPLPSGAMPQAOAMRLKENNFTV 723
DB 659 VVALGIG---LPMRRRI-VKRTLRLLQSEHVEPLTPSGEAPQALLRIKE----- 709
QY 724 SFMLRPVYSASHELYTKGIWIPDGENVKIPVAILKVLRENTSPKANEILDEAVYAGV 783
DB 710 TEFKKIKYLGSGAFGTGKGIWIPGEKVKIPVAILKELRENTSPKANEILDEAVYMASV 769
QY 784 GSPYVRLTGCTNSTVOLVQMPYGLLHVHNRGLSDODLLNCOMIAKMSYLE 843
DB 770 DNPVCRLLGLCTLSTVOLITQMPFGCLLYREHKNIGSQYLLNMCVOLAKMNTLE 829
QY 844 DVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHYADGQVPIKMALESILRR 903
DB 830 DRLVHRDLAARNVLYKTPQVVKITDFGLAKLIGAEKEHYHAEGKVPILKMALESILHR 889
QY 904 RFTHSDVWSIGVYTWELMTGAKPYDGIIPAREIPDLLEKGBRLPQPICTIDVYMIWK 963
DB 890 IYTHOSDWSIGVYTWELMTGSKPYDGIIPASEIISILKEGBRLPQPICTIDVYMIWK 949
QY 964 CMWISSECRPFRELVESEFSMARDPQRFVYIQ-NEGLPASPLDSTFYRSLLEDDMGD 1022
DB 950 CMWIDADSRPFRELVEFSMARDPQRYLYIQGDERHNLSPDTSNRYRPMDEDDMD 1009
QY 1023 LYDAEYLVPOQGFPCPDPAQAGMHHRRSSSTRSGGDLTLGLEFSEEBARSPLA 1082
DB 1010 VYDADEYLVPOQGF-----SSPSTRFTPL 1035
QY 1083 PSEGAGSVFPGDGLGMAKGLQSLPTDPSPLQRYSDPTVPUSRT--GGYVAPFLCS 1140
DB 1036 SLSKTSN--NSTVACITRNGLQSCPIKEDSFLQRYSDPTGALTEDSIDTFL----- 1087
QY 1141 PPEYVNPQDYRPPSPREGLPLPARPAGATLEBRKATLSFGKNGVVDVFAFGAVENP 1200
DB 1088 PVPEYINQ-SVKKRPAQSVQNPVYHNGDNLB-----APSRDPHYQD--PHSTAVGNP 1136
QY 1201 EYL-TPQGAARQPHRPPAPFARDNLVYMDQ-----DP-----PERGAPSTF 1243
DB 1137 EYLVNQ-----PFCVNSTFSDSPAHMAQKSHQISLDNPYQODFFPKAKPNGIF 1187
QY 1244 KGTPTAENPEYL 1255
DB 1188 KGS-TAENAEYL 1198

```

RESULT 5  
 A33183  
 epidermal growth factor receptor precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999  
 C:Accession: A33183; A43818; S24942; A28941; S45325; I49643  
 R:Luettecke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N  
 Genes Dev. 8, 359-413, 1994  
 A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor  
 A:Reference number: A33183; PMID:94170986; PMID:8125255  
 A:Accession: A33183  
 A:Molecule type: mRNA  
 A:Residues: 1-1210 <LUE>  
 A:Cross-references: GB:U03425

R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.  
 Oncogene 6, 673-676, 1991  
 A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site  
 A:Reference number: A43818, MUID:91232866; PMID:2030916  
 A:Accession: A43818  
 A:Molecule type: mRNA  
 A:Residues: 1-714 <AVI>  
 A:Cross-references: GB:X59698  
 R:Eslinger, D.P.; Serrero, G.  
 Submitted to the EMBL Data Library, June 1992  
 A:Reference number: S24942  
 A:Accession: S24942  
 A:Molecule type: mRNA  
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>  
 A:Cross-references: EMBL:Z12608  
 R:Heisermann, G.J.; Gili, G.N.  
 J. Biol. Chem. 263, 13152-13158, 1988  
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated  
 A:Reference number: A28941, MUID:88330814; PMID:3138233  
 A:Accession: A28941  
 A:Molecule type: protein  
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,  
 R:Hubbs, M.L.; Dunn, A.R.; Alexander, W.S.  
 Submitted to the EMBL Data Library, April 1994  
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor  
 A:Reference number: S45325  
 A:Accession: S45325  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-971, 'K', 973-1210 <VER>  
 A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CA55587.1; PID:G488831  
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993  
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse  
 A:Reference number: I49643; MUID:93126380; PMID:7678348  
 A:Accession: I49643  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 12-20, 22-132 <RES>  
 A:Cross-references: GB:I06864; NID:G193001; PIDN:AA53029.1; PID:G567201  
 C:Genetics:  
 A:Gene: EGFR  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphoprotein  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:648-670/Domain: transmembrane #status predicted <TM>  
 F:712-977/Domain: protein kinase homology <KIN>  
 F:720-728/Region: protein kinase ATP-binding motif  
 F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental  
 F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental  
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental  
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental  
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental

Query Match 44.3%; Score 3036; DB 2; Length 1210;  
 Best Local Similarity 48.6%; Pred. No. 8, 4e-118; Indels 118; Gaps 26;  
 Matches 622; Conservative 172; Mismatches 367;

244 CTGPHSDCLALPHNSGICELCPALVYNTDTESMNPBERITFGASCVTACPPNY 303  
 240 CTGPHSDCLVCCQFODEATCTKDCPCPLMYNPTTYQMDVNPBEKYSFGACVKKCPNY 299  
 304 LSTVGSCTVCPPLHNOEVTAEADGTGCEKSKPCAVCYGLGQVYIKANSKFIGITELE 363  
 300 VVTHGSCVRAKCGDYIEV-EDGIRCKCKDCGCRVKNIGIGERK-DLISINATNIK 357  
 358 HFVKCTAISGDLHLTPVAFKGDSTFRPPLDPRLEILTKYKEITGIFLLIQAMDNDMDL 417  
 423 SVFQNLQVIRRIILNAGYSTLTGGLGSMVGLSELSRELSGALLHHNHTLCVHTVPM 482  
 418 HAFENLEIRRTYOHQGFSLAVVGLNTSLGLSLKEISGDVYISGNRLCTANTINM 477  
 483 DQLEFNPQALHTANRPEDECEVGEGLACHOLCARHCWGPPTQCVNCSQFLRGQCEVE 542  
 478 KKLFGTPQKTKIMNMRBEKCKAVNHYCNPLCSGEGCWGEPRDVCSCQVNSRGRECE 537  
 543 ECRVYGLPREYVNAARHCLPHPECPQNGSVTCFGEADQVACAHYKDPFCVAPRS 602  
 538 KCMILEGPRERFVENSECICQHPCLPQAMNITGTGRPDNCCICAHYIDSPHCVKCTPA 597  
 603 GVKPDLSTYMPIWKFDEEGACQPCPINCSTHSCVDLDGCPAEQASPLTISVAVGIL 662  
 598 GIMGENNTL-VKRYADANNVCHLANCTYGCAGPLOGCEVWSGPMIPISLAIIGVGL 656  
 663 LVVLGVVFGI-LIKRQOKRKTYMRRLLOETLEVEPLTSGAMPNOQRILKEFNMF 721  
 657 LFIIV-VALGIGLEFRRRSHIVKRTLRLEORELVEPLTSGAMPNOQRILKEFNMF 711  
 722 TVSEFLARPKYSASHLETVYKGYIMPDENVKIPAIIVLEBNTPSKANKEILDEAYMA 781  
 712 -TEFKKRVUSGAGFTVYKGLMPEBEKXKIPVATIKELNLSIPKANKIILDEAYMA 769  
 782 GVGSFYVSRLLIGICTSTVOLVTLQMPYGLLDHYVNRGLSGQDLLNMCQIAKMSY 841  
 770 SVDNPHVCRLLGICLTSTVOLITQMPYGLLDHYVNRGLSGQDLLNMCQIAKMSY 829  
 842 LEDVYVNRDLAARVYLVKSPNHYKITDPGLARLLIDETRYHADGCVPIKMMALLESIL 901  
 830 LEDRVLVRDLAARVYLVKSPNHYKITDPGLARLLIDETRYHADGCVPIKMMALLESIL 889  
 902 RRRFTQSDWMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYIM 961  
 890 HRIYHQSDWMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPOPICTIDVYIM 949  
 962 VKCMWIDSECRPRFELVSEFSKMARDDQRFVIVQ-NEDDLQPASLSDTFYRSLLEDMD 1020  
 950 VKCMWIDSECRPRFELVSEFSKMARDDQRFVIVQ-NEDDLQPASLSDTFYRSLLEDMD 1009  
 1021 GDLVDAEYLVPOQGFCEPDPAKAGVHHRRHSSSTRSGGDLTLLESEEBEAPSP 1080  
 1010 EIVVDADAEYLVPOQGFCEPDPAKAGVHHRRHSSSTRSGGDLTLLESEEBEAPSP 1035  
 1081 LAPSEGASDVDFDGLGMAKGLQSLTTPBSPLOQRYSEDPVYVLPSET-DGVVAFLT 1138  
 1036 LLSLSLSTASN-NSTVACINRNSGCRVKEDAFQRYSSPPTGAVTNDN-IDAPL 1087  
 1139 GSPOPEYVNPQVROPSPSPREGLPAARPAAGATLERAKTSPGKNGVYKQVAFAGAVE 1198  
 1088 --PVEEYVNO-SVPRKPAQSONPYTHNQPLHP-----APGDLHYQN-PRSNVAG 1134  
 1199 NPEYV-TPQGAAPQPHPPAPSPAFDNLVYWDQ-----DP-----PERGAPPS 1241  
 1135 NPEYVNTAQ-----PTCLSSGFNSPALMIQSHQSMQSDMDNDYQODFFPKETKENG 1185  
 1242 TEKGTPTAENPEYVGLDVP 1260  
 1186 IFKG-PTAENAYLVKRAPE 1203

## RESULT 6

TVCHLV

epidermal growth factor receptor precursor - chicken

N/Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB

C/Species: Gallus gallus (chicken)

C/Date: 28-Feb-1986 #sequence\_revision 05-May-1995 #text\_change 04-Feb-2000

C/Accession: A27720; A00643

R/Lax, I.; Johnson, A.; Hawk, R.; Sap, J.; Bellot, F.; Winkler, M.; Ullrich, A.; Vennart

Mol. Cell. Biol. 8, 1970-1978, 1988

A/Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou

A/Reference number: A27720; MUID:88261272; PMID:3260329

A/Accession: A27720

A/Molecule type: mRNA

A/Residues: 1-1223 &lt;LAX&gt;

A/Cross-references: GB:M20386

R/Nilsen, T.M.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Citterenden, L.B.; Raines, M.

Cell 41, 719-726, 1985

A/Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro

A/Reference number: A00643; MUID:85228222; PMID:2988784

A/Accession: A00643

A/Molecule type: mRNA

A/Residues: 585-1223 &lt;NLT&gt;

A/Cross-references: GB:M10066

C/Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor

C:Specific protein kinase

F:1-30/Domain: signal sequence #status predicted &lt;SIG&gt;

F:31-1223/Product: epidermal growth factor receptor #status predicted &lt;MAT&gt;

F:31-654/Domain: extracellular #status predicted &lt;EXT&gt;

F:81-307/Domain: EGF receptor extracellular domain repeat &lt;EEL&gt;

F:397-610/Domain: EGF receptor extracellular domain repeat &lt;EE2&gt;

F:655-677/Domain: transmembrane #status predicted &lt;TM&gt;

F:678-1223/Domain: intracellular #status predicted &lt;INT&gt;

F:719-984/Domain: protein kinase homology &lt;KIN&gt;

F:727-735/Region: protein kinase ATP-binding motif

F:136-702/Binding site: carboxylate (Ser) (covalent) #status predicted

F:192-650/Binding site: carboxylate (Ser) (covalent) #status predicted

F:687/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted

F:754/Active site: Lys #status predicted

F:1100-1183/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match

Best Local Similarity 44.08; Score 3013.5; DB 1; Length 1223;

Matches 620; Conservative 177; Mismatches 354; Indels 153; Gaps 28;

QY 8 RMGLLALLPPGAA-----STOYCTGDMKRLPAPETHLDMRLHYGGCQVGNIE 61

DB 13 RGAALVLLLVALLGALCAVEKVCQSTNKKLTQLGHEVDFHFTSLQRMVNCVLSNLE 72

QY 62 LLYLPTNASLSFQDIOGVGYVLIHNVQVPLQRLIRVGTQLFEDNYALAVDNGD 121

DB 73 ILYVEHNRDLFTLKTIGEVAGYVLIANVDVIPLEMLQIRGVLYDNSFALAVSNYH 132

QY 122 PLNNTPTVTGASPGGRELRLSLTELKGVLIQRPOLCYODITIMQIFPKNNQL 181

DB 133 -MKTKQ-----GLNELPMKRLSELINGVKISNNPKLCNMDYLVANDIIDTSKR-PL 182

QY 182 TLID-TNRSRACHPCSPMCKGSRGWSESSDCSLTRTVAGGCA-RCKGPLPTDCHQ 239

DB 183 TVLDFASNLSSCPKCHNCTEDHGWGEGNQCTLTFTVICAQCGSRGCRKVPDSCCHQ 242

QY 240 CAAGCTGPKHSDCLACIHPHNSGICELHCPALVTYNTDTRESHPNREGRTFGASCYTAC 239

DB 243 CAAGCTGPRSDCLACKRFDDATCKTCCPLVLYNPTTYQMKVNBGKSFATCVRE 302

QY 300 PNYVLTSDVGSCTLVCLHNOEYTAEDGTQRCCKSPCARVCYGLGMQYIKANSKFIGI 359

DB 303 PHNYVLTDSGSCVRSNCTDYEV-EEHGVARKKCKKCGSLGKVCNGIGLKLKILS-INA 360

QY 360 TELF-PAGCKKIFGSLAFLESPFDGDPASNTAPLQPEQLQVFTLEBITGYLYISKMPDS 418

DB 361 TNIDSFKCTKINGVSVLIPVAFGDATFKTLPLDPKKLDVFRYTKELISGLILLQAMDND 420

QY 419 LPDLSPVQLOVIRGRILHNGAYSLTLOGLSWGLSLAEISGLLIHNTYLCFVN 478

DB 421 ATDLAFAENLEIRRTKHOGQYS-LAVNLKIOSGLSLKELISGDIAMKNKVLCYAD 480

QY 479 TVPMQQLFRNPHQALHTANRPEDECVGSLACHQLCRGKCGWGPFCVVCOSQFLRQ 538

DB 481 TNMMSLRATOSQKTKIIONNRKNDCTDRHVCDELCSDVCGMGSPFHCSGCRFFSNK 540

QY 539 ECVEECRYLQGLPREYVNAHRLCPHCEQPNQNG---SVTCFGEPAQCVAACHYKDPF 595

DB 541 ECVKQCNILQGEPRFEDSKCLPCHSECLVONSTAYNTCSGSPBDHCKAHFIDGPH 600

QY 596 CVARCPGKPDLSMPIMKPFDESGAQCPCINQTHSCVLDLDDGCAEQAASPLTSIV 655

DB 601 CVKACPAGVGENDTL-VMKYADANAVQLCPNCTRCCKGPGEGCP---NGSKTPSLA 656

QY 656 SAVV-GILLVVLGVVFGILLKRRQOKRKYTMRLLOETELVEBTLFSGAMPNOQMKRI 714

DB 657 AGVVGGLCLVVGIGLGLYLR-R-HYRKRTLRLLDEBELVEBTLFSGEAPNAHLRI 715

QY 715 LKEPNNFVSGFWLRPKYSASHLETYVYGMIPDGENYKIPATVYLAENSPKANKTEL 774

DB 716 LKE-----TEFKKKVLSGSAFGTVYVGLWIPSEKVKIPVAIKELREATSPKANKTEL 769

QY 775 DEAYVAGVSGVYVRLIGICLTSTVQLVTCMLPVGCLLDHRENRGRGSDLLNMCQ 834

DB 770 DEAYVAGVSDNHVRLIGICLTSTVQLVTCMLPVGCLLDYIREKNDIGSYLYLNKCVQ 829

QY 835 IAKGNSYLEDVRLVARDLAARVNVKSNHYKITDFFGLARLLIDETEVHADGKVPYIKW 894

DB 830 IAKGNNVLEERLVRDLAARVNVKTPQHVTITDFGLAKLGADEKRYHAEGKVPYIKW 889

QY 895 MLESILRRFTHQSDVWSYGTWELMTFGAKVPDGI-PAREIPDLLEKRELPQPPICT 954

DB 890 MLESILRRFTHQSDVWSYGTWELMTFGSKPDGI-PASEISSVLEKGERLPQPPICT 949

QY 955 IDVWIMVYKCMWIDECPPRFELVSESRNARDPQRFVQI-NEDLGPASLSTFFRS 1013

DB 950 IDVWIMVYKCMWIDDSKPKFRELIAESKVARDPRLVLVIGDERMHLPSPTDSKFFRT 1009

QY 1014 LLEDMDGDLVDAEYLYVPOGCFPCPDPAAGCAVHHRHSSSTRSGGDLTGLSE 1073

DB 1010 LMEEDMDIEDVADADYLYVPHOGF-----NSPST----- 1038

QY 1074 EEARPSPL-----ASEGASDVFPDGLGMAKGLQSLPTDPPFLQRYSEDPVPLPS 1128

DB 1039 ---SRPLLSLSLATSNNSATNCTD-----RNGQGHVPREDSFVQRYSSDPGNLE 1087

QY 1129 EY-DGYVAPLTCSPQPEYVQPDVPPQPSRREGPLPAPAPAGATLERAKTSLPGKNGV 1186

DB 1088 ESIDDGFL-----PAPEYVQ--LMPKPS-----TAMVQNCI 1118

QY 1187 VKDVF-----AFGAVNPEYVITPQGGAPQHPHPPASPAFNDNYNDQ 1231

DB 1119 YNNISLTAISKLPMDSRVONSHSTVANDPEYI-----NTNOSPLAKTVFESSPYMIQ 1170

QY 1232 -----DPEP-----RGAPPESTFGTPTAENPEYGLDVP 1260

DB 1171 SGNHQINLDNPYQDDPLPNEKPKRGLKVPAAENPEYLRVAAP 1214

## RESULT 7

A47253

epidermal growth factor receptor, HER4 - human

C/Species: Homo sapiens (man)

C/Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999

C/Accession: A47253

R/Florman, G.D.; Colouesco, J.M.; Whitney, G.S.; Green, J.M.; Carleton, G.W.; Foy, L.

Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993

A/Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide

A:Reference number: A47253; MUID:93189574; PMID:8383326  
 A:Accession: A47253  
 A:Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-1308 <P10>  
 A:Cross-references: GB:107868; NID:9337359; PID:AA859446.1; PID:9337360  
 A>Note: sequence extracted from NCBI backbone (NCBI:126842)  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; growth factor receptor; protein kinase homology  
 F:716-981/Domain: protein kinase homology <KIN>  
 F:724-732/Region: protein kinase ATP-binding motif

Query Match 42.1%; Score 2884.5; DB 2; Length 1308;  
 Best Local Similarity 44.2%; Pred. No. 1.5e-11; Indels 179; Gaps 29;  
 Matches 598; Conservative 185; Mismatches 390;

9 WGLLLALLPAGA---STOYCTGTDMKRLPASPTHLMRLHYOGCOVQGNELTY 64  
 8 WVVSLVLAAGTVQPSDSQSYCAGTENKLSLSDLEQOYALRKYENCCEVWAGNLEITS 67  
 65 LPTNASLFLDIOENGVYLAHNOYROYLQRLTYRGTOEFEDYALAVLDNDPLN 124  
 68 IENRDLSEFLRSVEVGVYVALLNQRVPLENLIRKGLIEDRYALATFLNTRKDG 127  
 125 NTPVGTASPGGLRELOLRSLTELKGVLIQRNPOLCYODTILMXDIFKNNQLATLI 184  
 128 NF-----GLQELGKMLTEILNGVYVDQMKFLCYADTIHMODIVANPMSNLTLV 178  
 185 DTNRSRACHPSCSPCKSRCKGSESDOQSLTRVYVAGGC-ARCKGLPTDCHEQCAAG 243  
 179 STNGSSGGRCHKSCGTG-RCWGPTEHRCQTLTRVCAEQDCGRYGVYVDCCHRECAAG 237  
 244 CTGPKHSDCLACLHFNHSGICELHCPALVTYNTDFESMPNPEGRYFGASCYACAPY 303  
 238 CSGKPDJDCFCACMFNSGACVTCQPTFYVNPPTFQLEHNFNAKYVGAFCVKKCPHF 237  
 304 LSTVGSCITVCPHNOETVAEDGTQRCCKSKACVYCGLMQYIKANSKRLGITELE 363  
 298 V-VDSSSCVRACPSSKMEV-BENGIRKCKPCTDLCPRACGIGISLMSAQYVDSNDK 355  
 364 FAGCKTFGLAFPESEPDGPANTAPLOPEQLOFETLEITGYLYISAMPDPLS 423  
 356 FINTKINGNIFLVTGHDGPAVAIEADEKLVNRYREITGFLINSGWPNMDF 415  
 424 VFQNLQVIRGILHNGAVSLTQGLGISMGLRSLRELSGLAIHNNHLCVHYVPMW 463  
 416 VFSNLVTIGGRVLSGLLILKQOGITSLQFOSLKEISAGNITYITNSNLVYHITMT 475  
 484 QLFENPHQALHTANRDECEVSGGLACHQLCARHONGGPGPOCVACQFLGQCEVEE 543  
 476 TLESTINORIVIRDRKAEKNTAAGWCNHLCSDDGCGEPDCLCRFSSRGRITIES 535  
 544 CRVLQGLPREVYNAHCLPCHPECP-ONGSVTCFGEADQVCAHAYKPPFCVAPCS 602  
 536 CNLYDEGEREFENSGICVECDPQCEKMEGDLTCHGDPDCTKCSHFQKGPVCEKCPD 595  
 603 GVRDLSMPIMKRPDEGACQPPINCTHSCVLLDKG-----PAQGRASP 651  
 596 GLQGANSE--LEKADPDRECHPCHPCTQCGNPSHDCIYYPWTHSTLPQHAR-TPL 652  
 652 TSISAVV-GILLVVLGVVFGILIKRQCKIRKYTRRLLOETELTEPTPSGAMPNOA 710  
 653 --IAAGVIGFLIVIGLTFAYVVRKRSIK-KKRALRRL-ELELVEPLPSTANQA 708  
 711 QMRILKEPNTVTFWMLRVKVSASHLETYYKGIWIDGENVKIPVAIKVLRENTSKAN 770  
 709 QLRILKE-----TELKRVKVLGAGATYVKGIMVEGETVKIPVAIKILNETGPKAN 762  
 771 KEIDEAIVMAGVSPYSRLGICITSTVOLVQLMPPYGLDHPVENGRGLSGQDILN 830  
 763 VEPFDEALIMASMPHPLVRLGVCLSTTOLVQLMPPHGLLEVEHNDNIGSQDILN 822  
 831 WCMQIAKMSYLEDRVLRVDRDLAARNVLVSPNHVKITTDGLARLIDIDETEHADGKV 890

DB 823 WCVQIAKGMATLEERLVRDLAARNVLVSPNHVKITTDGLARLIDIDETEHADGKM 882  
 QY 891 PIKMMALLESILRRPETHOSDVMSXGVYVWELMTFGAKPYDGI PAEIPDLLEKGRLLP 950  
 DB 883 PIKMMALLECITHYKETHOSDVMSXGVYVWELMTFGAKPYDGI PTREIPDLLEKGRLLP 942  
 QY 951 PICTIDVYIMYKCMWISSECPREPRRELVSERARARPPQFVYVQND-LGPASPLDT 1009  
 DB 943 PICTIDVYIMYKCMWISSECPREPRRELVSERARARPPQFVYVQND-LGPASPLDT 1002  
 QY 1010 FRSLLIEDMDGLVDAEYLVPCQGFCCPDPAFAGAGGVHHRSSSTRSGGDLTGL 1069  
 DB 1003 FQNLIDEDDLMDMADEYLVV-QAFNIPPP-----IYSRAIDENRS-----ELGH 1050  
 QY 1070 EPSEEBAPRS-----PLAF-SGASGVYVQDGLMGCA 1100  
 DB 1051 SPPPAYTPMSGNFYVRDGFPAEQGVSVYRAPRTSTIPBAFVAQATALEIFDSCNGT 1110  
 QY 1101 AKGLQSLPTHDSPLQRYSEPTVPLPS-----ETDGYVAPLTCSPQPEVYVQNDVP 1153  
 DB 1111 LKPYAPHYQEDSSTQRYSDPTVPAPEKSPGELDESGYVPMRDKRQETLVVE--- 1167  
 QY 1154 QPPSPREGEPLPARAGATLEERAKTLSPKNGVQVDFAFGAVENPEYLTPOGGAAP 1213  
 DB 1168 -----ENPFVSR-----KNGDLQ-----ALDNPETHNASNG----- 1194  
 QY 1214 HPEPA-----FSPAFNLVYVQDDPPPERGA- 1238  
 DB 1195 -PPKADEYVNEPLVNTFANTLGAEYVKNLISMPEKAKAFDNPDMYNSLPPRSTL 1253  
 QY 1239 -PPSTFKGTP-----AENPEYL 1255  
 DB 1254 QHPDYLGEXSTKYFKQNGRIRPIVAENPEYL 1285

## RESULT 8

S06142  
 protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish  
 N:Alternate names: epidermal growth factor receptor homolog; kinase-related transfc  
 C:Species: Xiphophorus maculatus (southern platyfish)  
 C:Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 18-Feb-2000  
 C:Accession: S06142; S13809  
 R:McIntosh, J.; Adam, D.; Malitschek, B.; Maeueller, W.; Raulf, F.; Telling, A.; Raulf, F.  
 A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing 1  
 A:Reference number: S06142; MUID:90015140; PMID:2797166  
 A:Accession: S06142  
 A:Molecule type: DNA  
 A:Residues: 1-1166 <W10>  
 A:Cross-references: EMBL:X16891; NID:965290; PID:CAA34770.1; PID:965291  
 R:Adam, D.; Maeueller, W.; Schartl, M.  
 Oncogene 6, 73-80, 1991  
 A:Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiph  
 A:Reference number: S13807; MUID:91125882; PMID:1846957  
 A:Accession: S13809  
 A:Status: preliminary; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 821-1025; 'N',1027-1098; 'A',1100-1166 <ADA>  
 A:Cross-references: EMBL:X56319; NID:965284; PID:CAA39763.1; PID:965285  
 C:Genetics: mrk  
 A:Map position: Y  
 A:Anticodon: 872/3; 947/1; 979/3; 1025/3; 1056/1  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein;  
 F:1-25/Domain: signal sequence #status predicted <SID>  
 F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>  
 F:707-972/Domain: protein kinase homology <KIN>  
 F:715-723/Region: protein kinase ATP-binding motif

Query Match 37.8%; Score 2592; DB 1; Length 1166;  
 Best Local Similarity 44.1%; Pred. No. 1.5e-99;





```

Db      238 PDUTDPCACHHNDGACVPCQPLVYNKLTFLQLEBNPTTKYQGVCAVASCPRNFV-V 296
Qy      307 DWGSCCTLVCPHNOETAEADGTQCEKSNPCARVGYIGMOYIKANSKF--IGITELE- 363
Db      297 DQNSCVACPPDKMEVD-KNGLKCEPCGGLCPKACEGTG-----SGSRFQYVDSNIDG 350
Qy      364 FACKXIFGSLAFLEPFDDPASNAPLQPELOQVETLEETITGLYLYISAMPDLSLDLS 423
Db      351 FVNCXKLGNDLITGLNDPMMKIPALDEKLVNFRVREITGLINQSMPPHNNFS 410
Qy      424 VFQNLQYIRGRILHNGAYS-LTLQGLGISWLGRLSRELGSGLALIHNNHLCFHTVPM 482
Db      411 VFSNLTITGGRSLYNRGFSLLIMKNLNVTLGFSRLSKETISAGRIYISANRQLCYHSLNM 470
Qy      483 DQLFNPHQALLHTA-NREDECVGEGLAQCLCAHCHGHPPTQCVCNSQPLRQGEV 541
Db      471 TKVLRGTEERLQIKNRRPRDCVAGKVDPLCSSGCGPFGQCLSRNRSRGVCV 530
Qy      542 EECRVLOGLPREYVNAHCLPCHPECCOPNGSVTCFEPADQCVACAHYKDPFCVARGP 601
Db      531 THCNFLNGEPRERFAHEAFCSCHPECCPMEGTATCNGSGSDTCAQCHFRDGHCVSSCP 590
Qy      602 SGVKPDLSTYMPWKFPPEBEGACQPCPINCTHSCVDLDDKCPABQRA---SPLTIVSA 657
Db      591 HGVLG--AKGPIKYRDPVQNECRCHENCCTQCKGPELOQLQTLVLTIGKTLTALTY 648
Qy      658 VVGLVAVVGVVFGILIKRROKIR-KYTRRLQTELEVEEPTSGAMPNOAMRIK 716
Db      649 IAG--LVVIFMLOGTFLYKRGRIQKRAMRYLEGESEIEFLDLS-EKANKVLAIRK 705
Qy      717 EFNNFTVSEFWLRVPKVSASHL-ETVYKGINIPQGENYKIPVAIKVIRENTSPANKELID 775
Db      706 ETE-----LRKIKVLGSGVFGTVHKGWIPGESIKIPVCKVIEDKSGROSFOAVLD 758
Qy      776 EAYVMAGVSPYVSRILGICLSTVOLVLTQMPYGCILDVRENRLGSGODLNNCMOI 835
Db      759 HMLAIGSLDHAHRYRLGLCPGSSSLQVLTQYLPGLSLDHYRQKRLGQLLNNGVQI 818
Qy      836 AKGMSYLEDEVRLVHRDLAARNVLYKSPNHVKITDFGIARLLDIDETRYHAGGKVP 895
Db      819 AKGMYLLEHGMVHRNLAARNVLLKSPQVADFGVADLLPPDDKQLTSEAKTPIKMM 878
Qy      896 ALBSILRRFTHODVWSGYVTWELMTFGAKPYDGI-PAREIDPLEKGRLOPPICT 955
Db      879 ALBSIFRGKYTHQSDVWSGYVTWELMTFGAEPYAGRLAEVPLLEKGERLQPOICTI 938
Qy      956 DVYMINVCKMVIDSECRPRRELVSFSSRMARDPQREVVITQNEDLGPA--SPUDSTFYR 1012
Db      939 DVYVMVWKCMVIDENIRPTEKELANFTRMARDPRYLVIKRES-GRGIAPGPEPHGLTN 997
Qy      1013 SLEDDDMGDLVNAEVLVPOQGFPCDPAPAGAGMHHNRSSSTSSGGGLTLGLEP- 1071
Db      998 KKEEVLLEPELDLIDLLEAED-----NLATTLTSSALSLPVGTL 1038
Qy      1072 SEEEAPRSPPLAPSEAGSDVFDGLGMAKGLQSLPTHD-PSPLQRYSEDPTVPLP- 1127
Db      1039 NRPRGSGSLSPSSGY-MPMNGSLGSCQESANSSGSERCPRVSLH-----PMRGC 1091
Qy      1128 ---SETDGYVA-----PLTCSPOPE---YVNOPIVRPQPPRRRGP- 1162
Db      1092 LASESEGHVTSAEALQEKVSMCRSRSPRFRGDSAYHSORHSLTLPVPLSPFGL 1151
Qy      1163 ---LPAARPAATLEERAKTSLF-GKNGV-----KQVFARGAIVENPEYLPQ 1206
Db      1152 EEDVNGVWPDTHLKGTSPSRBETLSSVGLSVLGEDEED-----EVEYNNRK 1203
Qy      1207 GGAAPQHPPPAPSPAFDNLTYND-----QDPPRGAPSTFTFGTPTAENPEY 1254
Db      1204 RHRSP-PHPRPSSLBEIAGEYMDVGSDLASLGSSTQSCPLHVPVIFTAGTTPDEDEY 1262
Qy      1255 L 1255
Db      1263 M 1263

```

```

RESULT 10
JC4387
Epidermal growth factor receptor homolog precursor - rat
N.Altimate names: ErbB3 protein; HER3 protein
C.Species: Rattus norvegicus (Norway rat)
C.Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C.Accession: J04387
R.Hellier, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A.Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant prot
A.Reference number: J04387; MUID:96096535; PMID:8522190
A.Accession: J04387
A.Molecule type: mRNA
A.Residues: 1-1339 <HEL>
A.Cross-References: GB:U29339; NID:G915389; PID:G915390
A.Experimental source: liver
A.Note: The authors translated the codon AAC for residue 369 as Thr and GTT for res
C.Comment: This protein is a functional heregulin receptor that transduces signals
C.Genetics:
A.Gene: ErbB3
C.Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C.Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prote
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>
F:640-659/Domain: transmembrane #status predicted <TM>
F:705-970/Domain: protein kinase #status predicted <KIN>
F:713-721/Region: protein kinase ATP-binding motif
F:939,1051,1156,1194,1219,1257,1259,1273,1286,1325/Binding site: phosphate (Ty
Query Match 32.8%; Score 2248.5; DB 2; Length 1339;
Best Local Similarity 39.8%; Pred. No. 2,2e-85;
Matches 514; Conservative 171; Mismatches 440; Indels 167; Gaps 37;
Qy      3 LAALRWGSLALLPPGA---STQVCTGTDMKRLRPSPEHIDMLHLYGGCCVQGN 59
Db      7 LQVLC---FLTLARGSEMNSQAVCGTNGLSVTGDADQYOTLYKVEKCEVYNGN 62
Qy      60 LETLTLPTNASLSPFQDIOEVGYVLIHANCYRQVPLRLRVRSTOLFEDNYVALAVLDN 119
Db      63 LEIVLTGNHADSFLQWIREVTAYLVAMNESVPLRLNLRVRSTQYVDGKFAIFVM-- 120
Qy      120 GDPPLNTPPTVTSAGSGLELQRLSLTEILKGVLIQKRPOLCYDITLWKDIFPKNOL 179
Db      121 ---LNYNT---NSHALROLKFTQLTLEILSGVYIEKDKYCHMDITDMRDIVR- 170
Qy      180 ALTLIDYRSRACHPGSPKSGRCMGSESSDCSLRTVCAAGC-ARCKGFLPTDCCHE 238
Db      171 GAEIVYKXNGANCPCPCHEVCKG-RKMGSPDDCQLITTTICAPQCNKRCFGENPQCHD 229
Qy      239 QCAAGCTGPKHSDCLALPHNSGICELHCPALVTYNTDTESMNPREGRYTFGASCYTA 298
Db      230 ECAGGCGSQDQDTCACRRFRFNSGACVRCPEPLVYNKLTFLQLEBNPTTKYQGVCAV 289
Qy      299 CPYNTLSTVSGCTLVCPHNOETAEADGTQCEKSNPCARVGYGL--GMQYIYANSKF 356
Db      290 CPNHFV-VQDTCVRACPDKMEVD-KHGLKACEPCGGLCPACGTSGSRYQTVDSN 347
Qy      357 IGITELEPAQCKKIFGSLAFLEPFDDPASNAPLQPELOQVETLEETITGLYLYISAMP 416
Db      348 ID----GFVNCXKILGNDLITGLNVPMKIPALDEKLVNFRVREITGLINQSMPP 403
Qy      417 DSLPDLSTYQNIQVIRGRILHNGAYS-LTLQGLGISWLGRLSRELGSGLALIHNNHLC 475
Db      404 PHMHNFSVFSNLTITGGRSLYNRGFSLLIMKNLNVTLGFSRLSKETISAGRIYISANRQL 463
Qy      476 FVHTVPMQLEFNPHQALLHTA-NREDECVGEGLAQCLCAHCHGHPPTQCVCNSQPLRQ 534
Db      464 YHSLNMTNRLRGPSEERLDIKYRDLPECLAEKGVCPPLSSGCGCWPRAPQGLSGANY 523
Qy      535 LFGQCEVECRVLOGLPREYVNAHCLPCHPECCOPNGSVTCFEPADQCVACAHYKDP 594

```



Db 524 SREGVYTHCNPLGSEPREFVHNAQCFSCHEPCLPMEGTSTVNGSGSDACARCAHPRDGE 583  
 QY 595 FCYARCPSSGVKPRDLSVYPIKFPREBEACQPCPINCNSHC--VDLDKXGAPAGQASPL 652  
 Db 584 HCVNSCPHGLG--AKSPIKYKPDQNECPCHENCTGGNGPELOCLGQAVLSKPH 641  
 QY 653 SIYSAVAVGILLVVLGVFGILIKRQOKIR-KYIMRRLQETELVEPLTPSGAMENQAO 711  
 Db 642 LVIAVYTG--LAVILMLGSGFLWGRRIQNKRAMRYLERGSEIPLDPS-EKANKVL 698  
 QY 712 MRLIKERNFTVSFWLVPKVSASHL-ETVYKGIWIDGENVKIPVAKYLRENTSFKAN 770  
 Db 699 ARIFKETE-----LRKLYLGGVFGVYHKGIMPEGESIKIPVCIKYIEKSGQSF 751  
 QY 771 KEILDEAVYVMAVGSFVYSRLIGICTSTVOLVTLQMPYGCILDHVENRGLSGODL 830  
 Db 752 QAVTDHMLAVOSLDHAIYRLGLGCPSSSLQVLYQYPLDGLDHVQNHETIGPOLLN 811  
 QY 831 WCMQIAKGMSTYEDVRLVHRDLAARNVLYKSPNHVKTITDGLARLIDIDETVHADGKV 890  
 Db 812 WGVQIAKGMVYLEHSHVHRDLARNVLMKSPSCQVADFGVADLPPDDKOLLHSEAKT 871  
 QY 891 PIKMMALSTLRREFTHOSDVMSYGVYVWELMTFGAKPYDGIAPREIPLDLEKGERLPOP 950  
 Db 872 PIKMMALSTLHFGKITHOSDVMSYGVYVWELMTFGAKPYDGIAPREIPLDLEKGERLPOP 931  
 QY 951 PICTIDVYMWKCMIDSECRPRFRELVSFMSARDPQRFVYIQNEDLGAPASPLDSTF 1010  
 Db 932 QICTIDVYMWKCMIDSECRPRFRELVSFMSARDPQRFVYIQNEDLGAPASPLDSTF 988  
 QY 1011 YRSLIEDDDMDLVDAAEYVLPVQGFPCPDPAFAGMWHHRHSSSTRSGGDLTIGLE 1070  
 Db 989 EPSVLTTEL-----QEALELEPL-----DLDDLE 1014  
 QY 1071 PSEF-----EAPRSLAPSEG-----AGSDVEPDGDLGMAAK 1102  
 Db 1015 AEEBGLATSGSALSLPTGLTLTRPGSGSLSPSGVPMNQSLGEACLSAYLGREQ 1074  
 QY 1103 GLGSLPTHPDSEPLQRYSEDPTVPLPSETDGYV---APL-----TC-----SPOPE-- 1144  
 Db 1075 FSRPISLH-PIPRGR-----PASESEGHVTSSEALQEKVSVCSRSRSPRFRGCD 1126  
 QY 1145 --VYNQDPRVROPSPREGR-----LPAPAPGATLEBAKTLSP-GKKGVV-- 1187  
 Db 1127 SAYHSQHSLLTPVTPSPGLEBEDGNGYVMPDTHLRGASSSREGTLSSVGLTE 1186  
 QY 1188 --XDVFAFGAVENPEYLTPOGGAAPQHP 1216  
 Db 1187 EDED-----EEYEMNRKRKRGSP-PRPP 1209

RESULT 11  
 TVFVTV  
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus  
 N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase  
 C:Species: avian leukosis virus, ALV  
 C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 11-Jun-1999  
 C:Accession: B00643; A00643  
 R:Nilsen, I.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M.  
 Cell 41, 719-726, 1985  
 A>Title: c-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro  
 A:Reference number: A00643; MUID:85282822; PMID:2988784  
 A:Molecule type: mRNA  
 A:Residues: 1-658 <NT>  
 A:Cross-references: GB:M10066; GB:M13981; NID:g211749; PIDN:AAA4763.1; PID:g211750  
 A>Note: in Genbank entry CHKRBEP, release 109.0, the source is designated as Gallus gal  
 C:Comment: This protein is synthesized as a gag-env-erbB protein.  
 C:Genetics:  
 A:Gene: gag-env-erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P  
 P1-6/Product: gag protein (fragment) #status predicted <GAG>

F17-59/Product: env protein (fragment) #status predicted <ENV>  
 F160-698/Product: protein-tyrosine kinase erdB #status predicted <ERB>  
 F194-459/Domain: protein kinase homology <KID>  
 F202-210/Region: protein kinase ATP-binding motif  
 F1229/Active site: Lys #status predicted

Query Match 24.6%; Score 1685.5; DB 1; Length 698;  
 Best Local Similarity 50.1%; Pred. No. 1.6e-62;  
 Matches 362; Conservative 84; Mismatches 145; Indels 131; Gaps 19;

QY 578 GPEADQVCANHYKDPPEVCARCPGSGVPRDLSYMPIMKRPDEBACQPCPINCNSCVL 637  
 Db 60 GP--DHCMKCAHFDGPHCVKACPAGVAGENLT--VWKADANACQLCHPCTROCKRP 116  
 QY 638 DDKGCPAEORASPLTSYSAVY-GILLVVLGVFGILIKRQOKIRKYIMRRLQETEL 696  
 Db 117 GLEGCP--NGSKTSPSIAAGVGGILLCLVAVGIGLYLRR-HIVRKTLRLRLQEREL 172  
 QY 697 VERPLTPSGAMPQAOIKRIKERNFTVSFWLVPKVSASHLETVYKGIWIDGENVKIPV 756  
 Db 173 VERPLTPSGAPQAHRLIKR-----TEFKVYKLVGSAFGVYKGLMIPGEEKVYIPV 226  
 QY 757 AIKVLRENTSPANKELIDEAVYVMAVGSFVYSRLIGICTSTVOLVTLQMPYGCILDHY 816  
 Db 227 AIKELRENTSPANKELIDEAVYVMAVGSFVYSRLIGICTSTVOLVTLQMPYGCILDHY 286  
 QY 817 RENGRLSGQDLINCMQIAKGMSTYEDVRLVHRDLAARNVLYKSPNHVKTITDGLARL 876  
 Db 287 REHKDNGISQYLLNMCVQIAKGMVYLEHSHVHRDLAARNVLYKTPQHAKITDFGAKLL 346  
 QY 877 DIDETEYHADGGVPIKMMALSTLRREFTHOSDVMSYGVYVWELMTFGAKPYDGIAPRE 936  
 Db 347 GADKEKHYHAGGVPIKMMALSTLRITTHOSDVMSYGVYVWELMTFGSKPYDGIAPRE 406  
 QY 937 IPDLLEKGERLPOPPICTIDVYMWKCMIDSECRPRFRELVSFMSARDPQRFVYIQ 996  
 Db 407 ISSVLEKGERLPPPICTIDVYMWKCMIDSECRPRFRELVSFMSARDPQRFVYIQ 466  
 QY 997 -NEDLCPASPLDSTFRSLIEDDDMDLVDAAEYVLPVQGFPCPDPAFAGMWHHRHRS 1055  
 Db 467 GDRKMLPSPSTSKFRTLMEEEDMDLVDADYVLPVQGF-----NS 510  
 QY 1056 SSTRSGGDLTIGLEBSESEAPRPL-----APSEAGSDVEPDGDLGMAKGLQSLPTH 1110  
 Db 511 PSL-----SRPLSLSSAHSNNNSAINCID-----RNGQGHPR 544  
 QY 1111 DPSPLQRYSEDPTVPLPSET--DGYVAPLTCSPQRYVNOQDVRPOPSPREGRPLPAAP 1168  
 Db 545 EDSFVORYSSDPTGNFLEESIDDFL-----PAPRYVQ--LMPKKFS----- 585  
 QY 1169 AGATLERAKTLSPGKXGVYVDF-----AFGAVENPEYLTPOGGAAPQ 1213  
 Db 586 -----TANVQOQIYNNISLTISKLPMSRIONSGHSTVADNPEY-----NT 627  
 QY 1214 HPPAPSPAFDNLVYWDQ-----DPEP-----RGAPSTFGKTPAENPEYLGID 1258  
 Db 628 NQSPPLAKTVESPPYIQSGNHQINDNPYQODPLPNETKPNGLLKVPAAENPEYLRVA 687  
 QY 1259 VP 1260  
 Db 688 AP 689

RESULT 12  
 TVYTH  
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain  
 C:Species: avian erythroblastosis virus  
 C>Date: 18-Apr-1984 #sequence\_revision 18-Apr-1984 #text\_change 11-Jun-1999  
 C:Accession: A00644; A38022  
 R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.  
 Cell 35, 71-78, 1983  
 A>Title: The erbB gene of avian erythroblastosis virus is a member of the src gene  
 A:Reference number: A00644; MUID:84026539; PMID:6313229

A:Accession: A00644  
 A:Molecule type: DNA  
 A:Residues: 1-604 <YAM>  
 A:Cross-references: GB:K01216; NID:G209676; PIDN:AAA42400.1; PID:G209678  
 A:Database: B. Henry, C. Benalissa, M. Biserte, G. Claverie, J.M. Saule, S. Martin, Science 224, 1456-1459, 1994  
 A:Title: Sequencing the erbB gene of avian erythroblastosis virus reveals a new type of  
 A:Reference number: A38022; MUID:64223957; PMID:6328658  
 A:Accession: A38022  
 A:Molecule type: DNA  
 A:Residues: 1-28, 'W', 30-139, 'F', 141-145, 'V', 147-152 <DEB>  
 A:Cross-references: GB:K02006  
 C:Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific  
 F:130-395/Domain: protein kinase homology <KIN>  
 F:138-146/Region: protein kinase ATP-binding motif  
 F:165/Active site: Lys #status predicted

Query Match 23.7%; Score 1622; DB 1; Length 604;  
 Best Local Similarity 50.0%; Pred. No. 5.8e-60;  
 Matches 348; Conservative 80; Mismatches 136; Indels 132; Gaps 17;  
 QY 587 CAHYKDPFCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVLDLDDKGPAGQ 646  
 DB 3 CAHFDGPHCVKACPAVGLGENDTL-VKXVADANAVALQCHPCTRCCKGPGEGCP--- 58  
 QY 647 RASPLTIVSAVV-GILLVVLGVFGILLKRRQKIRKTYMRLLDQETLVEPLTPSGA 705  
 DB 59 NGSKTSPSIAGVGGGLCLVAVGLGILYLR-R-HYRKRTLRLLQERLVEPLTPSGE 117  
 QY 706 MPNQAQRILKEFNFTVSFMLRVPKVSASHLETYVKGIMIPGENVKIIVALKVIRENT 765  
 DB 118 APNQAHRILKE-----TEFKKVKVLGSGAFGTITKGLMIPSEKXKIVALKELREAT 171  
 QY 766 SPKANKILDEAYMAGVSPYVSRLLGICLTSTVQLVQLMRYGCLLDHRENGRLGS 825  
 DB 172 SPKANKILDEAYMASVNDPHVCRLGLICLTSTVQLITQIMPGCLLDYIRHKONIGS 221  
 QY 826 QDLNMCQIAKMSYLEDVRLVHDLAARNAVLYKSPNHKITDFFGLARLLDDETEYHA 885  
 DB 232 QYLLNMCVQIAKMSYLEERLVRDLAARNAVLYKTPQHKITDFFGLAKLGGADEYHA 291  
 QY 886 DGGVPIKMALESILRRFTHOSDVSYGVYVWELMTFSKAPYDGIIPAREIDPLEKGE 945  
 DB 292 EGCKVPIKMALESILRRITHOSDVSYGVYVWELMTFSKAPYDGIIPAREIDPLEKGE 351  
 QY 946 RLQPPICITDVIYIMKCMWIDSECRPRELVSFSKARDPQRFVYIQ-NEDLGPAS 1004  
 DB 352 RLQPPICITDVIYIMKCMWIDSECRPRELVSFSKARDPQRFVYIQ-NEDLGPAS 411  
 QY 1005 PLDSTFYRSLDEDDMDGLVDAEYVLPQGFPCPDPAAGAGVHHRRHSSSTRSGGD 1064  
 DB 412 PTDSKFYRTLMEEDMEDIVDADEYLVPHQGF-----NSPST----- 449  
 QY 1065 LTUGLESEBEAPRSP-----APSEGASGVDFGDLGMAKGLQSLPHHDSPLQRY 1119  
 DB 450 -----SPTPLSLISATSNNSATNCID-----RNGQGHPRVDSFVQRY 489  
 QY 1120 EDTPVPLPSET--DGYVAPLTCSPQPEYVNOQDVPRQPPSPREGPLPAARPAATLERAK 1177  
 DB 490 SDPTGNFLSEIDDGFL-----PAPEYVNO--LMPKRPSTAM----- 524  
 QY 1178 TLSPGKNGVYKDVAF-----GAVENPERLTQGGAAPOHPPPA 1218  
 DB 525 -----VQNTINFLSLTAISKLPMDSRYQNSHTAVDNPEYL-----NTQSP 568  
 QY 1219 FSPAFLNLYWDDPPEPGAPPTFGKPTAENPEY 1254  
 DB 569 AKTVFESSPYWIGSGNQ-----INLDNDY 594

RESULT 13  
 S35745  
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus  
 C:Species: avian erythroblastosis virus  
 C:Date: 03-Mar-1994 #sequence\_revision 26-May-1995 #text\_change 28-Feb-1997  
 C:Accession: S35745  
 R:Yennstrom, B  
 Submitted to the EMBL Data Library, March 1993  
 A:Reference number: S35743  
 A:Accession: S35745  
 A:Molecule type: DNA  
 A:Residues: 1-544 <VEN>  
 A:Cross-references: EMBL:X12707  
 C:Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-speci  
 F:135-400/Domain: protein kinase homology <KIN>  
 F:143-151/Region: protein kinase ATP-binding motif  
 F:170/Active site: Lys #status predicted

Query Match 22.9%; Score 1569; DB 2; Length 544;  
 Best Local Similarity 52.7%; Pred. No. 7.9e-58;  
 Matches 334; Conservative 73; Mismatches 129; Indels 98; Gaps 16;  
 QY 578 GREADCVACAHYKDPFCVACPSGVKPDLSYMPIMKFPDEGACQPCPINCTHSCVLD 637  
 DB 1 GP--DHMKCAHFDGPHCVKACPAVGLGENDTL-VKXVADANAVALQCHPCTRCCKGPG 57  
 QY 638 DDKGCPAEGRASPLTIVSAVV-GILLVVLGVFGILLKRRQKIRKTYMRLLDQETL 696  
 DB 58 GLEGGP---NGSKTSPSIAGVGGGLCLVAVGLGILYLR-R-HYRKRTLRLLQERL 113  
 QY 697 VEPLTPSGAMPNQAQRILKEFNFTVSFMLRVPKVSASHLETYVKGIMIPGENVKIIV 756  
 DB 114 VEPLTPSGAMPNQAHRILKE-----TEFKKVKVLGSGAFGTITKGLMIPSEKXKIV 167  
 QY 757 AIKVLRENTSPKANKILDEAYMAGVSPYVSRLLGICLTSTVQLVQLMRYGCLLDH 816  
 DB 168 AIKVLRENTSPKANKILDEAYMASVNDPHVCRLGLICLTSTVQLITQIMPGCLLDY 227  
 QY 817 RENGRLGQDLNMCQIAKMSYLEDVRLVHDLAARNAVLYKSPNHKITDFFGLARLL 876  
 DB 228 RENGRLGQYLLNMCVQIAKMSYLEERLVRDLAARNAVLYKTPQHKITDFFGLAKL 287  
 QY 877 DIDETEYHAGGKVPKIMMALESILRRFTHOSDVSYGVYVWELMTFSKAPYDGIIP 936  
 DB 288 GADKEXYHAGGKVPKIMMALESILRRITHOSDVSYGVYVWELMTFSKAPYDGIIP 347  
 QY 937 IDDLLEKGERLQPPICITDVIYIMKCMWIDSECRPRELVSFSKARDPQRFVYIQ 996  
 DB 348 ISSVLEKGERLQPPICITDVIYIMKCMWIDSECRPRELVSFSKARDPQRFVYIQ 407  
 QY 997 -NEDLGPASPLDSTFYRSLDEDDMDGLVDAEYVLPQGFPCPDPAAGAGVHHRRH 1055  
 DB 408 GDERHNLBSPPTSKYRTLMEEDMEDIVDADEYLVPHQGF-----NS 451  
 QY 1056 SSTSGGGDGLUGLESEBEAPRSP-----APSEGASGVDFGDLGMAKGLQSLPH 1110  
 DB 452 PST-----SPTPLSLISATSNNSATNCIDRNG-----H 481  
 QY 1111 DSPPLQRYSEDPVPLPSETGQVAPLTCSPQPEYVNOQDVPRQPPSPREGPLPAARPA 1170  
 DB 482 -----PYREDGFL-----PAPEYVNO--LMPKRPSTAMONQLYNIS 517  
 QY 1171 ATLERAKTLSPGKNGVYKDVAFAGAVENPEYL 1203  
 DB 518 LTAISKLPIDSRYN-----SHSTAVDNPEYL 544

RESULT 14  
 S00727  
 kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi

C:Species: avian erythroblastosis virus  
 C:Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 28-Feb-1997  
 C:Accession: S00727  
 R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.  
 Oncogene Res. 1, 265-278, 1987  
 A>Title: Common site of mutation in the erbB gene of avian erythroblastosis virus mutant  
 A:Reference number: S00727; MUID:88217326; PMID:2897102  
 A:Accession: S00727  
 A:Molecule type: DNA  
 A:Residues: 1-545 <SCO>  
 A:Cross-references: EMBL:X06943  
 C:Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; phosphotransferase  
 F:135-400/Domain: protein kinase homology <Kin>  
 F:143-151/Region: protein kinase ATP-binding motif

Query Match 22.8%; Score 1562; DB 2; Length 545;  
 Best Local Similarity 52.7%; Pred. No. 1.5e-57;  
 Matches 334; Conservative 72; Mismatches 130; Indels 98; Gaps 16;

QY 578 GPEADQVCAHYKPPFCVAPCPGKVDLSYMPDKRPDEGACQCPINCTHSCVDL 637  
 DB 1 GP--DHCKCAFIIGPHCVKACPAVGLGENDTL-VKXADANAVCQLCHPCTKCKGP 57  
 QY 638 DDKGPAPGASPLTISVAVV-GILLVVLGVFGLIKRQOKIRKYMRLLOETEL 696  
 DB 58 GLEGCP---NSKSTPSIAAGVVGGLLVVGLGILYLRR-HIVRKRLRLLOEREL 113  
 QY 697 VEPLTPSGAMPQAOAKRIKERNNTVSWLVPKVASHLETYVKGWIDGKWKIPV 756  
 DB 114 VEPLTPSGAMPQAOAKRIKERNNTVSWLVPKVASHLETYVKGWIDGKWKIPV 167  
 QY 757 AIKVLRENTSPKANKELIDEAYVMAGVSPYVRLIGLICTSTVOLTLQMPYGLDHY 816  
 DB 166 AIKELRENTSPKANKELIDEAYVMAGVSPYVRLIGLICTSTVOLTLQMPYGLDHY 227  
 QY 817 RENGRLSQDLINCMQIAKMSYLEVRLVHRLAARLVKSPNHVKTITDGLARLL 876  
 DB 228 RENGRLSQDLINCMQIAKMSYLEVRLVHRLAARLVKSPNHVKTITDGLARLL 287  
 QY 877 DIDETEVADGSKVYIKMMALLESILRRFTHSDVMSYGVWELMTGAKPYGIPARE 936  
 DB 288 GADEKEVYHAGSKVYIKMMALLESILRRFTHSDVMSYGVWELMTGAKPYGIPARE 347  
 QY 937 IPDLLEKGERLPQPICTIDVYIMVKCMTIDSECRPPRELVSFSKMAADPORFVVIQ 996  
 DB 348 ISSVLEKGERLPQPICTIDVYIMVKCMTIDSECRPPRELVSFSKMAADPORFVVIQ 407  
 QY 997 -NEDLGPAFLDSTFYRSLDDDDMDGLVDAEYLVPOGFCPPRAGAGVHHRRRS 1055  
 DB 408 GDERHMLSPDTSKRYRLTMEEDMEDVDADEYLVPOGFCPPRAGAGVHHRRRS 451  
 QY 1056 SSTRSGGDDLTLGLEPSEEARSPD-----APSEAGSDVFDGDLGMAKGLSLPTFH 1110  
 DB 452 PST-----SRPLLSLSLSTNSNATNCIDRNG-----H 481  
 QY 1111 DPSPLOKRYSEDTVLPSETQGYVAPLTCSPQPEVYNQPDVRRPQPSRREGPLPAAPAG 1170  
 DB 482 -----PVREDFL-----PAPEVYNQ-LMPKKPSTAMVQNOIYVYIS 517  
 QY 1171 AT-TERAKTLSPGKNVYKVDVAFGAVENEPEYL 1203  
 DB 518 LTAISKLPMDSKYQN-----SHSTAVDNPEYL 544

RESULT 15  
 GQFF  
 Epidermal growth factor receptor - fruit fly (*Drosophila melanogaster*)  
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erbB  
 C:Species: *Drosophila melanogaster*  
 C:Date: 17-Mar-1987 #sequence\_revision 17-Mar-1987 #text\_change 11-Jun-1999

C:Accession: A00640; A38021  
 R:Lynne, E.; Glazer, L.; Segal, D.; Schlessinger, J.; Shilo, B.Z.  
 Cell 40, 559-607, 1985  
 A>Title: The *Drosophila* EGF receptor gene homolog: conservation of both hormone bi  
 A:Reference number: A00640; MUID:85124611; PMID:2982499  
 A:Accession: A00640  
 A:Molecule type: DNA  
 A:Residues: 1-1330 <LIV>  
 A:Cross-references: EMBL:X03054  
 R:Madaworth, S.C.; Vincent III, W.S.; Bloddeau-Wentworth, D.  
 Nature 314, 178-180, 1985  
 A>Title: A *Drosophila* genomic sequence with homology to human epidermal growth fac  
 A:Reference number: A38021; MUID:85137938; PMID:2983222  
 A:Accession: A38021  
 A:Molecule type: DNA  
 A:Residues: 1-832-866, 'V', 868-943, 'QTPSLVK' <MAD>  
 A:Cross-references: EMBL:X02293; NID:97922; PID:CA26157.1; PID:97929565  
 C:Comment: This sequence is tentative because the introns have not been identified  
 C:Genetics:  
 A:Gene: FlyBase:Egfr  
 A:Cross-references: FlyBase:FBgn0003731  
 A:Map position: 2.57p  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; p  
 F:1-732/Domain: extracellular #status predicted <EXT>  
 F:733-764/Domain: transmembrane #status predicted <TM>  
 F:765-1330/Domain: intracellular #status predicted <INT>  
 F:808-1072/Domain: protein kinase homology <Kin>  
 F:816-824/Region: protein kinase ATP-binding motif  
 F:122,300,324,363,518,668,695,700/Binding site: carbohydrate (Asn) (covalent) #sta  
 F:774/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predi  
 F:843/Active site: Lys #status predicted  
 F:1181/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #status p

Query Match 22.7%; Score 1557.5; DB 1; Length 1330;  
 Best Local Similarity 29.0%; Pred. No. 5.e-57;  
 Matches 404; Conservative 178; Mismatches 419; Indels 391; Gaps 42;

QY 80 VQGVYLAHQVQVPLQRLRIVRGTLF-----EDNYALAVLDNGDPLNNTTPVGTASP 134  
 DB 38 ITNYIVIGLDLIPETLSRYLQIRGRILFSLVSEKAYLFV-----TY 81  
 QY 135 GGLRELDRLTELKSGVLIQRNPQLCYDTITMKDIPIKNNOLATLIDNRSRACHP 194  
 DB 82 SKMTTLEIPDLRDLNQLQVGFHNHNYMLCHRRITOWSEIVSNGDAYNYDFTAPRECPK 141  
 QY 195 CSPYCKSRRCMGESSEDCQSLTRTVACAGCA--RCKGRLPTDCCHEGCAAGCTGPKXSDC 252  
 DB 142 CHECTHG-CWGEGRKNCQFSLKLTGSPQACAGRCYGRPKRCHLFCAGCGTGTQKDC 200  
 QY 253 LACLHNHSGICELHCPALVYNTDTFESMPNPEGRTTFGASCVTACPYNYLSTDVSGT 312  
 DB 201 IACNPFDEAVSKEECPMKRYNPPTVLETNPEGKAYAGATCVKECP-GHLRLDNGACV 259  
 QY 313 LVCPRLNQEYTAADGTGRCCKSPKAPVYGVGMQYIKANSKFIGITEL-----EFAG 366  
 DB 260 RSCPDQMDKGE-----CYPCNGPCKTC-----PGTVLHAGNDSFRN 300  
 QY 367 CKKIFGSLAFSPESFDG--DPAANTA-----PLOGEQLQVFTLEITGYLYISAMPDS 418  
 DB 301 CTVIDGNIRILDQTFSGQVYVANYMGPRYIPLDEBRRVFTVWEITGYLYIEGHPQ 360  
 QY 419 LPPLSVFQNIQVIRGLIHNGAY-SLTQGLGISMGLSRLSGGLAIHNTHLCPV 477  
 DB 361 FRLSYFRNLETTHGQLWESMAALAIYKSLYSLEMRNLKQISSGSVVIQNRDLQVY 420  
 QY 478 HTVPMDQLFRNPQALHTANRDEDC----- 504  
 DB 421 SNIRWPAIQKEPQKQVAVENLRADLGKFTLLISGVCHNIWHIPIAKREKNNHLGSV 480  
 QY 505 ----- 504  
 DB 481 QRGRLGSMHGSVPLQELQFWLHRLRLMYIQVINSIDPKSNEHQLTDACYSPSVPT 540

QY 505 -----VG 506  
Db 541 SLTIERARYAIOAGLAMELEQITARSASMRHSKTLPAEGROVPRWFLGVACASARAGIA 600  
QY 507 EGIA-----CHOLCARGHGWPGPTOCVNCSCPLRGCEVEECRVLOGLPREVY---N 556  
Db 601 EPLAGAVCRKCHPLCELTCTNYGHEQVCSKCTHYKRQCETEC-----PADHYTDEE 654  
QY 557 ARHCLPCEPCOPONGSVTCFGEBADQVACAHYK-----DPPF-----CVARCPG 603  
Db 655 QRECFQRHPEC---NG---CTGPGADDCSKCRNFLELDANETGPYVNSTMFNCTSKCPLE 708  
QY 604 VK-PDLSTYMPKWFPEEGACQCPINCHSCVDLDKGCFAEGRASPLTISVASAVGIL 662  
Db 709 MRHVNQYTAIGPY-----CAASPRSSKITAND-----VNIIFITIGAV 749  
QY 663 LVVVLGVFEGI-LIKRROOKIRRYT--MRLLQETELVEPLTSGAMPNOQMRLIKEFN 719  
Db 750 LVPTICILCVTYICROKOKAKKETVQMTALSGRESEPLRPSNIGANLCKRIYKDAE 809  
QY 720 NTFVSFWLAVPKV-SASHLETYKGIWIDGENVKIPVAIKVLRNTPSKANKETLDEAY 778  
Db 810 -----LRKGGVLGAGARVYKGVWPEGENVKIPVAIKELKSTGAESESEPLREAY 862  
QY 779 VMAGVSSPYVSRLLGICLTSTVQVLTQMLPYGCLLDHYRENGRLGSDLLNMCMQIAKG 838  
Db 863 IMASEEHVNLKLLAVCMSSQWMLITQMLPLGCLLDYVRNNRDKIGSKALNWSQIAKG 922  
QY 839 NSYLEDEVRLVHRDLAARNLVK---SPNHVKITDGLARLLDIDETEHADGKVPYKMM 895  
Db 923 NSYLEBEKRLVHRDLAARNLVRLLAGEDH---DFGLAKLLSSDSNEYKPAAGKMPKMWL 978  
QY 896 ALESILRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPPOPICTI 955  
Db 979 ALECIARNRVFTSKSDVMAFGVIMELTFCGRPHENIPAKDIPDLIEVGLKLEQPEICSL 1038  
QY 956 DVMYMWKCMYIDSECRPRFRELIVSEFSMARDPQRFVVIQNEIDLG--PASPLDSTFYRS 1013  
Db 1039 DIVCTLLSCMHLDAAMRPTFKQLTVFAEPARDPGRYALILGDKFTRLPA-----YTS 1091  
QY 1014 LLEDD---DMGDLVDAEYLVPOQGFPCDPAPAGAGMWHRRSSSTRSGGDLTLGLE 1070  
Db 1092 QBEKDLIRKLAFTTDSSEALAKPDDYLOPKALGFS-----HRTDCT----- 1133  
QY 1071 PSEEEAP-----RSLPAPSEAGASDVFDG---DLGMAKAGLQSLPTHDPSPLOYSED 1121  
Db 1134 ---DEMPKLNRYCKDPSNNSSTGDDERDSSAREVGNTLR----- 1171  
QY 1122 PTVPLPSEITDGVAPLTGSPQPEYVQPDVRPOPPSPREGPLPAARPAATLERAKITLSP 1181  
Db 1172 --LDLPVEDDDYIMP-TCQPGPNNNNMMA-----NPNONNMAAVGVAAGYV----- 1214  
QY 1182 GKNGVVVDVAFAGGAVENPEYL---TPQGAAPOPH-----PPPAF 1219  
Db 1215 -----DLIGVAVSVDNPEYLLNAQTLGVGESPIPTQITGIPVWGPGTMEVKVPMFGS 1267  
QY 1220 SP-APDNLTYMD 1230  
Db 1268 EPTSSDHEYND 1279

Search completed: July 22, 2003, 09:27:59  
Job time : 31.3921 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.6928 Seconds

(without alignments)  
4891.279 Million cell updates/sec

Title: SEQ4-695-709-14  
Perfect score: 6853  
Sequence: 1 MEALALCRWGLLALLPGA.....TFKGTPTAENPEYGLDVPV 1261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6641	96.9	1255	1 ERB2_HUMAN	P04626 homo sapien
2	5850	85.4	1257	1 ERB2_RAT	P06494 rattus norv
3	5834.5	85.1	1254	1 ERB2_MESAU	P06053 mesocricetu
4	3063	44.7	1210	1 EGFR_HUMAN	P00633 homo sapien
5	3037	44.3	1210	1 EGFR_MOUSE	Q01279 mus sapien
6	2884.5	42.1	1308	1 ERB4_HUMAN	Q15303 mus sapien
7	2865	41.8	1308	1 ERB4_RAT	Q62956 rattus norv
8	2608.5	38.1	1167	1 XMRK_XIPMA	P13388 xiphophorus
9	2343.5	34.2	1342	1 ERB3_HUMAN	P21860 homo sapien
10	2271.5	33.1	1339	1 ERB3_RAT	Q62799 rattus norv
11	1879	27.4	1426	1 EGFR_DROME	P04412 drosophila
12	1668.5	24.3	634	1 ERBB_AIV	P00534 avian leuko
13	1622	23.7	604	1 ERBB_AVIER	P00635 avian eryth
14	1595	22.3	703	1 EGFR_CHICK	P13387 gallus gall
15	1552	22.6	540	1 ERBB_AVIEU	P11373 avian eryth
16	1241	18.1	1323	1 LT23_CAEEL	P24448 caenorhabdi
17	1142.5	16.7	245	1 ERB2_MOUSE	P70424 mus musculu
18	700	10.2	1363	1 ILBR_BRALA	Q02466 brachyosco
19	670	9.8	1363	1 INSR_RAT	P15127 rattus norv
20	669.5	9.8	1372	1 INSR_MOUSE	P15208 mus musculu
21	665	9.7	1300	1 INSR_MOUSE	Q9WC14 mus musculu
22	665	9.7	1382	1 INSR_HUMAN	P06213 homo sapien
23	656	9.6	1297	1 IRR_HUMAN	P14616 homo sapien
24	655	9.6	1477	1 HTK7_HYDAT	Q25197 hydra atten
25	653	9.5	1607	1 MIPR_LYMSM	Q25410 lymaea stea
26	651.5	9.5	1300	1 IRR_CANPO	P14617 cavia porce
27	596	8.7	1367	1 IGR_HUMAN	P08069 homo sapien
28	590	8.6	1390	1 INSR_AEDAE	Q93105 aedes aegypt
29	582	8.5	2146	1 INSR_DROME	P09208 drosophila
30	578	8.4	1373	1 IGR_MOUSE	Q60751 mus musculu
31	574.5	8.4	1370	1 IGR_RAT	P24062 rattus norv
32	574	8.4	1114	1 RET_HUMAN	P07949 homo sapien
33	571	8.3	987	1 EPB4_HUMAN	P54760 homo sapien

34	569.5	8.3	984	1 EPB1_CHICK	O07494 gallus gall
35	566.5	8.3	984	1 EPB1_RAT	P09759 rattus norv
36	562.5	8.2	977	1 EPB2_MOUSE	Q03145 mus musculu
37	560.5	8.2	984	1 EPB1_HUMAN	P54762 homo sapien
38	558	8.1	902	1 EPBB_XENLA	Q91736 xenopus lae
39	556.5	8.1	976	1 EPB2_HUMAN	Q29317 homo sapien
40	551	8.0	987	1 EPB4_MOUSE	P54761 mus musculu
41	550.5	8.0	985	1 EPB4_XENLA	Q91571 xenopus lae
42	544	7.9	1068	1 EPB1_XENLA	Q91738 xenopus lae
43	539.5	7.9	1053	1 FAK1_CHICK	O08944 gallus gall
44	537	7.8	757	1 HT16_HYDAT	P53356 hydra atten
45	532.5	7.8	1052	1 FAK1_MOUSE	P34152 mus musculu

## ALIGNMENTS

```

RESULT 1
ID      ERB2_HUMAN          STANDARD;          PRT; 1255 AA.
AC      P04626;
DT      13-AUG-1987 (Rel. 05, Last sequence update)
DI      13-AUG-1987 (Rel. 05, Last sequence update)
DI      13-0UN-2002 (Rel. 41, Last annotation update)
DE      Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE      (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE      surface receptor HER2) (MIM 19).
GN      ERBB2 OR HER2 OR NGL OR NEU.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX      NCBI_TaxId=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=66118663; PubMed=3003577;
RA      Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA      Saito T., Toyoshima K.;
RT      "Similarity of protein encoded by the human c-erbB-2 gene to
RT      epidermal growth factor receptor.";
RL      Nature 319:230-234(1986).
[2]
[3]
[4]
[5]
[6]
[7]
[8]
[9]
[10]
[11]
[12]
[13]
[14]
[15]
[16]
[17]
[18]
[19]
[20]
[21]
[22]
[23]
[24]
[25]
[26]
[27]
[28]
[29]
[30]
[31]
[32]
[33]
[34]
[35]
[36]
[37]
[38]
[39]
[40]
[41]
[42]
[43]
[44]
[45]
[46]
[47]
[48]
[49]
[50]
[51]
[52]
[53]
[54]
[55]
[56]
[57]
[58]
[59]
[60]
[61]
[62]
[63]
[64]
[65]
[66]
[67]
[68]
[69]
[70]
[71]
[72]
[73]
[74]
[75]
[76]
[77]
[78]
[79]
[80]
[81]
[82]
[83]
[84]
[85]
[86]
[87]
[88]
[89]
[90]
[91]
[92]
[93]
[94]
[95]
[96]
[97]
[98]
[99]
[100]
[101]
[102]
[103]
[104]
[105]
[106]
[107]
[108]
[109]
[110]
[111]
[112]
[113]
[114]
[115]
[116]
[117]
[118]
[119]
[120]
[121]
[122]
[123]
[124]
[125]
[126]
[127]
[128]
[129]
[130]
[131]
[132]
[133]
[134]
[135]
[136]
[137]
[138]
[139]
[140]
[141]
[142]
[143]
[144]
[145]
[146]
[147]
[148]
[149]
[150]
[151]
[152]
[153]
[154]
[155]
[156]
[157]
[158]
[159]
[160]
[161]
[162]
[163]
[164]
[165]
[166]
[167]
[168]
[169]
[170]
[171]
[172]
[173]
[174]
[175]
[176]
[177]
[178]
[179]
[180]
[181]
[182]
[183]
[184]
[185]
[186]
[187]
[188]
[189]
[190]
[191]
[192]
[193]
[194]
[195]
[196]
[197]
[198]
[199]
[200]
[201]
[202]
[203]
[204]
[205]
[206]
[207]
[208]
[209]
[210]
[211]
[212]
[213]
[214]
[215]
[216]
[217]
[218]
[219]
[220]
[221]
[222]
[223]
[224]
[225]
[226]
[227]
[228]
[229]
[230]
[231]
[232]
[233]
[234]
[235]
[236]
[237]
[238]
[239]
[240]
[241]
[242]
[243]
[244]
[245]
[246]
[247]
[248]
[249]
[250]
[251]
[252]
[253]
[254]
[255]
[256]
[257]
[258]
[259]
[260]
[261]
[262]
[263]
[264]
[265]
[266]
[267]
[268]
[269]
[270]
[271]
[272]
[273]
[274]
[275]
[276]
[277]
[278]
[279]
[280]
[281]
[282]
[283]
[284]
[285]
[286]
[287]
[288]
[289]
[290]
[291]
[292]
[293]
[294]
[295]
[296]
[297]
[298]
[299]
[300]
[301]
[302]
[303]
[304]
[305]
[306]
[307]
[308]
[309]
[310]
[311]
[312]
[313]
[314]
[315]
[316]
[317]
[318]
[319]
[320]
[321]
[322]
[323]
[324]
[325]
[326]
[327]
[328]
[329]
[330]
[331]
[332]
[333]
[334]
[335]
[336]
[337]
[338]
[339]
[340]
[341]
[342]
[343]
[344]
[345]
[346]
[347]
[348]
[349]
[350]
[351]
[352]
[353]
[354]
[355]
[356]
[357]
[358]
[359]
[360]
[361]
[362]
[363]
[364]
[365]
[366]
[367]
[368]
[369]
[370]
[371]
[372]
[373]
[374]
[375]
[376]
[377]
[378]
[379]
[380]
[381]
[382]
[383]
[384]
[385]
[386]
[387]
[388]
[389]
[390]
[391]
[392]
[393]
[394]
[395]
[396]
[397]
[398]
[399]
[400]
[401]
[402]
[403]
[404]
[405]
[406]
[407]
[408]
[409]
[410]
[411]
[412]
[413]
[414]
[415]
[416]
[417]
[418]
[419]
[420]
[421]
[422]
[423]
[424]
[425]
[426]
[427]
[428]
[429]
[430]
[431]
[432]
[433]
[434]
[435]
[436]
[437]
[438]
[439]
[440]
[441]
[442]
[443]
[444]
[445]
[446]
[447]
[448]
[449]
[450]
[451]
[452]
[453]
[454]
[455]
[456]
[457]
[458]
[459]
[460]
[461]
[462]
[463]
[464]
[465]
[466]
[467]
[468]
[469]
[470]
[471]
[472]
[473]
[474]
[475]
[476]
[477]
[478]
[479]
[480]
[481]
[482]
[483]
[484]
[485]
[486]
[487]
[488]
[489]
[490]
[491]
[492]
[493]
[494]
[495]
[496]
[497]
[498]
[499]
[500]
[501]
[502]
[503]
[504]
[505]
[506]
[507]
[508]
[509]
[510]
[511]
[512]
[513]
[514]
[515]
[516]
[517]
[518]
[519]
[520]
[521]
[522]
[523]
[524]
[525]
[526]
[527]
[528]
[529]
[530]
[531]
[532]
[533]
[534]
[535]
[536]
[537]
[538]
[539]
[540]
[541]
[542]
[543]
[544]
[545]
[546]
[547]
[548]
[549]
[550]
[551]
[552]
[553]
[554]
[555]
[556]
[557]
[558]
[559]
[560]
[561]
[562]
[563]
[564]
[565]
[566]
[567]
[568]
[569]
[570]
[571]
[572]
[573]
[574]
[575]
[576]
[577]
[578]
[579]
[580]
[581]
[582]
[583]
[584]
[585]
[586]
[587]
[588]
[589]
[590]
[591]
[592]
[593]
[594]
[595]
[596]
[597]
[598]
[599]
[600]
[601]
[602]
[603]
[604]
[605]
[606]
[607]
[608]
[609]
[610]
[611]
[612]
[613]
[614]
[615]
[616]
[617]
[618]
[619]
[620]
[621]
[622]
[623]
[624]
[625]
[626]
[627]
[628]
[629]
[630]
[631]
[632]
[633]
[634]
[635]
[636]
[637]
[638]
[639]
[640]
[641]
[642]
[643]
[644]
[645]
[646]
[647]
[648]
[649]
[650]
[651]
[652]
[653]
[654]
[655]
[656]
[657]
[658]
[659]
[660]
[661]
[662]
[663]
[664]
[665]
[666]
[667]
[668]
[669]
[670]
[671]
[672]
[673]
[674]
[675]
[676]
[677]
[678]
[679]
[680]
[681]
[682]
[683]
[684]
[685]
[686]
[687]
[688]
[689]
[690]
[691]
[692]
[693]
[694]
[695]
[696]
[697]
[698]
[699]
[700]
[701]
[702]
[703]
[704]
[705]
[706]
[707]
[708]
[709]
[710]
[711]
[712]
[713]
[714]
[715]
[716]
[717]
[718]
[719]
[720]
[721]
[722]
[723]
[724]
[725]
[726]
[727]
[728]
[729]
[730]
[731]
[732]
[733]
[734]
[735]
[736]
[737]
[738]
[739]
[740]
[741]
[742]
[743]
[744]
[745]
[746]
[747]
[748]
[749]
[750]
[751]
[752]
[753]
[754]
[755]
[756]
[757]
[758]
[759]
[760]
[761]
[762]
[763]
[764]
[765]
[766]
[767]
[768]
[769]
[770]
[771]
[772]
[773]
[774]
[775]
[776]
[777]
[778]
[779]
[780]
[781]
[782]
[783]
[784]
[785]
[786]
[787]
[788]
[789]
[790]
[791]
[792]
[793]
[794]
[795]
[796]
[797]
[798]
[799]
[800]
[801]
[802]
[803]
[804]
[805]
[806]
[807]
[808]
[809]
[810]
[811]
[812]
[813]
[814]
[815]
[816]
[817]
[818]
[819]
[820]
[821]
[822]
[823]
[824]
[825]
[826]
[827]
[828]
[829]
[830]
[831]
[832]
[833]
[834]
[835]
[836]
[837]
[838]
[839]
[840]
[841]
[842]
[843]
[844]
[845]
[846]
[847]
[848]
[849]
[850]
[851]
[852]
[853]
[854]
[855]
[856]
[857]
[858]
[859]
[860]
[861]
[862]
[863]
[864]
[865]
[866]
[867]
[868]
[869]
[870]
[871]
[872]
[873]
[874]
[875]
[876]
[877]
[878]
[879]
[880]
[881]
[882]
[883]
[884]
[885]
[886]
[887]
[888]
[889]
[890]
[891]
[892]
[893]
[894]
[895]
[896]
[897]
[898]
[899]
[900]
[901]
[902]
[903]
[904]
[905]
[906]
[907]
[908]
[909]
[910]
[911]
[912]
[913]
[914]
[915]
[916]
[917]
[918]
[919]
[920]
[921]
[922]
[923]
[924]
[925]
[926]
[927]
[928]
[929]
[930]
[931]
[932]
[933]
[934]
[935]
[936]
[937]
[938]
[939]
[940]
[941]
[942]
[943]
[944]
[945]
[946]
[947]
[948]
[949]
[950]
[951]
[952]
[953]
[954]
[955]
[956]
[957]
[958]
[959]
[960]
[961]
[962]
[963]
[964]
[965]
[966]
[967]
[968]
[969]
[970]
[971]
[972]
[973]
[974]
[975]
[976]
[977]
[978]
[979]
[980]
[981]
[982]
[983]
[984]
[985]
[986]
[987]
[988]
[989]
[990]
[991]
[992]
[993]
[994]
[995]
[996]
[997]
[998]
[999]
[1000]

```

CC -1- P.TM. LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY).  
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN  
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
 CC OF 0.762; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL; M1767; AAA35808.1; JOINED.  
 DR EMBL; M1761; AAA35808.1; JOINED.  
 DR EMBL; M1762; AAA35808.1; JOINED.  
 DR EMBL; M1763; AAA35808.1; JOINED.  
 DR EMBL; M1764; AAA35808.1; JOINED.  
 DR EMBL; M1765; AAA35808.1; JOINED.  
 DR EMBL; M1766; AAA35808.1; JOINED.  
 DR EMBL; M1730; AAA35808.1; JOINED.  
 DR EMBL; M12036; AAA35978.1; -.  
 DR EMBL; X03363; CAA27060.1; -.  
 DR PIR; A25491; A25491.  
 DR PIR; A24571; A24571.  
 DR HSSP; P11362; IFGK.  
 DR GeneW; HGNC:3450; ERBB2.  
 DR MIM; 164870; -.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR InterPro; IPR004019; YLP\_motif.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR Pfam; PF02757; YLP; 2.  
 DR Prodom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; TYKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Polymorphism.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 1235 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 653 675 POTENTIAL.  
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 720 987 PROTEIN KINASE.  
 FT NE\_BIND 726 734 ATP (BY SIMILARITY).  
 FT BINDING 753 753 ATP (BY SIMILARITY).  
 FT ACT\_SITE 845 845 BY SIMILARITY.  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 220 227 BY SIMILARITY.  
 FT DISULFID 224 235 BY SIMILARITY.  
 FT DISULFID 236 244 BY SIMILARITY.  
 FT DISULFID 240 252 BY SIMILARITY.  
 FT DISULFID 255 264 BY SIMILARITY.  
 FT DISULFID 268 295 BY SIMILARITY.  
 FT DISULFID 299 311 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 511 520 BY SIMILARITY.  
 FT DISULFID 515 528 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 567 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 BY SIMILARITY.  
 FT MOD\_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARIANT 654 654 I -> V.  
 FT VARIANT 655 655 /FTID=VAR\_004077.  
 FT VARIANT 655 655 I -> V.  
 FT CONFLICT 1170 1170 /FTID=VAR\_004078.  
 FT SEQUENCE 1255 AA; 137909 MW; 3959DFDA04DC962 CRC64;  
 SQ  
 Query Match 96.9%; Score 6641; DB 1; Length 1255;  
 Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;  
 1 MELALQRMGLLALLPPGAASVQVCGTGMKRLPASPETHLMLRLHYGCGVQVQNTL 60  
 1 MELALQRMGLLALLPPGAASVQVCGTGMKRLPASPETHLMLRLHYGCGVQVQNTL 60  
 61 ELTYLPTNASISFLQDIQEVGYVLIANQVQVPLQRLIRVGTQFEDNYALAVLDNG 120  
 61 ELTYLPTNASISFLQDIQEVGYVLIANQVQVPLQRLIRVGTQFEDNYALAVLDNG 120  
 121 DPLNNTTPVYASGRRLQRLSITLILKGVLIQGNPOLCYVDITLMDIFKHNQGLA 180  
 121 DPLNNTTPVYASGRRLQRLSITLILKGVLIQGNPOLCYVDITLMDIFKHNQGLA 180  
 121 DPLNNTTPVYASGRRLQRLSITLILKGVLIQGNPOLCYVDITLMDIFKHNQGLA 180  
 121 DPLNNTTPVYASGRRLQRLSITLILKGVLIQGNPOLCYVDITLMDIFKHNQGLA 180  
 181 LTLIDITNRSRACHPCSPMKCKSRGCESSSEDCQSLRTVCAAGCARCKGRLPTDCHEQC 240  
 181 LTLIDITNRSRACHPCSPMKCKSRGCESSSEDCQSLRTVCAAGCARCKGRLPTDCHEQC 240  
 181 LTLIDITNRSRACHPCSPMKCKSRGCESSSEDCQSLRTVCAAGCARCKGRLPTDCHEQC 240  
 181 LTLIDITNRSRACHPCSPMKCKSRGCESSSEDCQSLRTVCAAGCARCKGRLPTDCHEQC 240  
 241 AAGCTGPRHSQCLALHPNHSGLICELHPALVYNTDTFESMPNPEGRYTFGASCVTACP 300  
 241 AAGCTGPRHSQCLALHPNHSGLICELHPALVYNTDTFESMPNPEGRYTFGASCVTACP 300  
 241 AAGCTGPRHSQCLALHPNHSGLICELHPALVYNTDTFESMPNPEGRYTFGASCVTACP 300  
 241 AAGCTGPRHSQCLALHPNHSGLICELHPALVYNTDTFESMPNPEGRYTFGASCVTACP 300  
 301 YNYLSTVGSCTLVCPPLHNOEVTAEADGTCRCCKSCARVCYGLGMQYIKANSKFTIGIT 360  
 301 YNYLSTVGSCTLVCPPLHNOEVTAEADGTCRCCKSCARVCYGLGMQYIKANSKFTIGIT 360  
 301 YNYLSTVGSCTLVCPPLHNOEVTAEADGTCRCCKSCARVCYGLGMQYIKANSKFTIGIT 360  
 301 YNYLSTVGSCTLVCPPLHNOEVTAEADGTCRCCKSCARVCYGLGMQYIKANSKFTIGIT 360  
 361 ELEPRAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLOVFEETLEITGYLYISAMPDSLP 420  
 361 ELEPRAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLOVFEETLEITGYLYISAMPDSLP 420  
 361 ELEPRAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLOVFEETLEITGYLYISAMPDSLP 420  
 361 ELEPRAGCKKIFGSLAFPESEFDGPASNTAPLQPEQLOVFEETLEITGYLYISAMPDSLP 420  
 421 DLSYFQNLQVIRGRILHNGAYSLTLOGIGISMLGRSIREGSGALIHNTHTLCFYHTV 480  
 421 DLSYFQNLQVIRGRILHNGAYSLTLOGIGISMLGRSIREGSGALIHNTHTLCFYHTV 480  
 421 DLSYFQNLQVIRGRILHNGAYSLTLOGIGISMLGRSIREGSGALIHNTHTLCFYHTV 480  
 421 DLSYFQNLQVIRGRILHNGAYSLTLOGIGISMLGRSIREGSGALIHNTHTLCFYHTV 480  
 481 PWDOLFRRPHQALHTANRPEDECVGSLACHQCARGHCGPPTCQVNCVSQFLRQEC 540  
 481 PWDOLFRRPHQALHTANRPEDECVGSLACHQCARGHCGPPTCQVNCVSQFLRQEC 540  
 481 PWDOLFRRPHQALHTANRPEDECVGSLACHQCARGHCGPPTCQVNCVSQFLRQEC 540  
 481 PWDOLFRRPHQALHTANRPEDECVGSLACHQCARGHCGPPTCQVNCVSQFLRQEC 540  
 541 VEECRVLOGLFREYVNAHCHPCHPPEQCPQNGSVTCGPREDDQCVACHYDPPFCVARG 600  
 541 VEECRVLOGLFREYVNAHCHPCHPPEQCPQNGSVTCGPREDDQCVACHYDPPFCVARG 600  
 541 VEECRVLOGLFREYVNAHCHPCHPPEQCPQNGSVTCGPREDDQCVACHYDPPFCVARG 600  
 541 VEECRVLOGLFREYVNAHCHPCHPPEQCPQNGSVTCGPREDDQCVACHYDPPFCVARG 600  
 601 PSQVYKPDLSYMPPIKFPDEEGACQPCPINTHSCVDLDDKQCPAEPAPLTSIVSAVG 660  
 601 PSQVYKPDLSYMPPIKFPDEEGACQPCPINTHSCVDLDDKQCPAEPAPLTSIVSAVG 660  
 601 PSQVYKPDLSYMPPIKFPDEEGACQPCPINTHSCVDLDDKQCPAEPAPLTSIVSAVG 660  
 601 PSQVYKPDLSYMPPIKFPDEEGACQPCPINTHSCVDLDDKQCPAEPAPLTSIVSAVG 660  
 661 ILLVYVGLVVGILLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAOMRLIKEFN 720  
 661 ILLVYVGLVVGILLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAOMRLIKEFN 720  
 661 ILLVYVGLVVGILLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAOMRLIKEFN 720  
 661 ILLVYVGLVVGILLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAOMRLIKEFN 720

QY 721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKIPVATKVLRENTSPKANELLIDEAVV 779  
 DB 720 -----LRKVKVLSGAFGTVTKGIWIPGENVKIPVATKVLRENTSPKANELLIDEAVV 773  
 QY 780 MAGVGSPPYVSRLLIGLITSTVQVLTVMPEYGLLDHVRNRRGLSQDLLNMCWQIAKGM 839  
 DB 774 MAGVGSPPYVSRLLIGLITSTVQVLTVMPEYGLLDHVRNRRGLSQDLLNMCWQIAKGM 833  
 QY 840 SYLEDVRLVRLDAAANNVLKSPNHVKITDFGLARLLIDETEHYHADGKVPILKMALES 899  
 DB 834 SYLEDVRLVRLDAAANNVLKSPNHVKITDFGLARLLIDETEHYHADGKVPILKMALES 893  
 QY 900 ILRRFTHQSDVWSYGVTVWELMTFGAKFYDGIIPAREIPDLLEKGERLPPQPICTIDVYM 959  
 DB 894 ILRRFTHQSDVWSYGVTVWELMTFGAKFYDGIIPAREIPDLLEKGERLPPQPICTIDVYM 953  
 QY 960 IMVKCMWIDSECPRRRELVSSESRMAPDQRFVVIQNDLGPASPLDSTFYRSLIEDDD 1019  
 DB 954 IMVKCMWIDSECPRRRELVSSESRMAPDQRFVVIQNDLGPASPLDSTFYRSLIEDDD 1013  
 QY 1020 MEDLVDAEELVYQGFPCPDPAFGAGVWHRHRSSTRSGGDLTLGLPSESEAPRS 1079  
 DB 1014 MEDLVDAEELVYQGFPCPDPAFGAGVWHRHRSSTRSGGDLTLGLPSESEAPRS 1073  
 QY 1080 PLAPSEAGSDVFDGGLGCAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTC 1139  
 DB 1074 PLAPSEAGSDVFDGGLGCAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYVAPLTC 1133  
 QY 1140 SPQPEYVNPQDVRPQPPSPREGPLPARPAGATLEPAKTLSPKXNVVMDVAFGAVEN 1199  
 DB 1134 SPQPEYVNPQDVRPQPPSPREGPLPARPAGATLEPAKTLSPKXNVVMDVAFGAVEN 1193  
 QY 1200 PEYLTPQGGAPAPPHPPPAFSPAFLNLYYWDQDPPERGAPSTFTFKGTPTAENPEYLGLDV 1259  
 DB 1194 PEYLTPQGGAPAPPHPPPAFSPAFLNLYYWDQDPPERGAPSTFTFKGTPTAENPEYLGLDV 1253  
 QY 1260 PV 1261  
 DB 1254 PV 1255

RESULT 2  
 ERB2\_RAT STANDARD; PRT: 1257 AA.  
 ID P06494;  
 AC 01-JAN-1988 (Rel. 06, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Epidermal growth factor receptor-related protein).  
 GN ERB2 OR NEU.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Neuroblastoma;  
 RX MEDLINE=86118662; PubMed=3945311;  
 RA Bargmann C.I., Hung M.-C., Weinberg R.A.;  
 RT "The neu oncogene encodes an epidermal growth factor receptor-related protein."  
 RL Nature 319:226-230(1986).  
 RN [2]  
 RP SEQUENCE OF 852-905 FROM N.A.  
 RC TISSUE=Sciatic nerve;  
 RX MEDLINE=91222560; PubMed=2025425;  
 RA Lai C., Lemke G.;  
 RT "An extended family of protein-tyrosine kinase genes differentially  
 RT expressed in the vertebrate nervous system."  
 RL Neuron 6:691-704(1991).

RN [3]  
 RP STRUCTURE BY NMR OF 650-668.  
 RX MEDLINE=92155181; PubMed=1346763;  
 RA Gullick W.J., Bottomley A.C., Lofis F.J., Doak D.G., Mulvey D.,  
 RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;  
 RT "Three dimensional structure of the transmembrane region of the proto-  
 RT oncogenic and oncogenic forms of the neu protein."  
 RL EMBO J. 11:43-48(1992).  
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.  
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A  
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
 CC ALPHA AND AMPHIREGULIN.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERB2 RECEPTORS.  
 CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: X03362; CA27059.1; ALT\_INIT.  
 CC PIR: A24362; TVRTNU.  
 CC HSP: P11362; IFGK.  
 CC InterPro: IPR000494; EGFR\_L\_domain.  
 CC InterPro: IPR000719; Euk\_Pkinase.  
 CC InterPro: IPR002174; Furin-like.  
 CC InterPro: IPR001245; Tyr\_Pkinase.  
 CC InterPro: IPR004019; YLP\_motif.  
 CC Pfam: PF00069; Pkinase; 1.  
 CC Pfam: PF00757; Furin-like; 1.  
 CC Pfam: PF01030; Recep\_L\_domain; 2.  
 CC Pfam: PF02757; YLP; 2.  
 CC ProDom: PD000001; Euk\_Pkinase; 1.  
 CC SMART: SM00261; FU; 3.  
 CC SMART, SM00219; TYKIC; 1.  
 CC PROSITE, PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC PROSITE, PS00109; PROTEIN\_KINASE\_TYR; 1.  
 CC PROSITE, PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 CC Proto-oncogene; Disease mutation.  
 CC SIGNAL 1 21 POTENTIAL.  
 CC CHAIN 22 1257 RECEPTOR PROTEIN-TYROSINE KINASE ERB2-2.  
 CC DOMAIN 22 654 EXTRACELLULAR (POTENTIAL).  
 CC TRANSMEM 655 677 POTENTIAL.  
 CC DOMAIN 678 1257 CYTOPLASMIC (POTENTIAL).  
 CC DOMAIN 159 369 CYS-RICH.  
 CC DOMAIN 473 646 CYS-RICH.  
 CC NP\_BIND 722 989 PROTEIN KINASE.  
 CC BINDING 728 736 ATP (BY SIMILARITY).  
 CC ACT\_SITE 847 847 ATP (BY SIMILARITY).  
 CC DISULFID 196 205 BY SIMILARITY.  
 CC DISULFID 200 213 BY SIMILARITY.  
 CC DISULFID 221 228 BY SIMILARITY.  
 CC DISULFID 225 236 BY SIMILARITY.  
 CC DISULFID 237 245 BY SIMILARITY.  
 CC DISULFID 241 253 BY SIMILARITY.  
 CC DISULFID 256 265 BY SIMILARITY.  
 CC DISULFID 269 296 BY SIMILARITY.  
 CC DISULFID 300 312 BY SIMILARITY.  
 CC DISULFID 316 332 BY SIMILARITY.  
 CC DISULFID 335 339 BY SIMILARITY.  
 CC DISULFID 513 522 BY SIMILARITY.

FT DISULFID 517 530 BY SIMILARITY.  
 FT DISULFID 533 542 BY SIMILARITY.  
 FT DISULFID 546 562 BY SIMILARITY.  
 FT DISULFID 565 578 BY SIMILARITY.  
 FT DISULFID 569 586 BY SIMILARITY.  
 FT DISULFID 589 598 BY SIMILARITY.  
 FT DISULFID 602 625 BY SIMILARITY.  
 FT DISULFID 628 636 BY SIMILARITY.  
 FT DISULFID 632 644 BY SIMILARITY.  
 FT MOD RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEU).  
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 85.4%; Score 5850; DB 1; Length 1257;  
 Best Local Similarity 85.8%; Pred. No. 3.2e-306; Index 10; Caps 4;  
 Matches 1084; Conservative 52; Mismatches 118; Indels 10; Gaps 4;

QY 1 MELAALCRMGILLALPPGAATQVCTGTDMLRLPASPEHLDMRLHYOGQVQGNL 60  
 DB 1 MELAALCRMGILLALPPGAATQVCTGTDMLRLPASPEHLDMRLHYOGQVQGNL 60  
 QY 61 ELTYLPNALSIFQDIQEVGYVLIHNOVROVPLORLIVAGTQIFEDVNAVLNDG 120  
 DB 61 ELTYLPNALSIFQDIQEVGYVLIHNOVROVPLORLIVAGTQIFEDVNAVLNDG 120  
 QY 121 DPLNNTPTVT-GASPGGLRELOLRLTEILKGVLIQORNPOLCYODTIIMKDFHKNOL 179  
 DB 121 DPLNNTPTVT-GASPGGLRELOLRLTEILKGVLIQORNPOLCYODTIIMKDFHKNOL 179  
 QY 121 DPQDNVAASIPRGTREBELRLQLRLTEILKGVLIQORNPOLCYODTIIMKDFHKNOL 180  
 DB 121 DPQDNVAASIPRGTREBELRLQLRLTEILKGVLIQORNPOLCYODTIIMKDFHKNOL 180  
 QY 180 ALTLIDTNRSPRACPCSPWCKSGRCWGESSEDCQSLTRIVACAGCARCKPLPTDCHEQ 239  
 DB 180 ALTLIDTNRSPRACPCSPWCKSGRCWGESSEDCQSLTRIVACAGCARCKPLPTDCHEQ 239  
 QY 181 APVDIDTNRSPRACPCSPWCKSGRCWGESSEDCQSLTRIVACAGCARCKPLPTDCHEQ 240  
 DB 181 APVDIDTNRSPRACPCSPWCKSGRCWGESSEDCQSLTRIVACAGCARCKPLPTDCHEQ 240  
 QY 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTESNPNEGRTFGASCVTAC 299  
 DB 240 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTESNPNEGRTFGASCVTAC 299  
 QY 241 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTESNPNEGRTFGASCVTAC 300  
 DB 241 CAAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTESNPNEGRTFGASCVTAC 300  
 QY 300 PNYLSTDVSGCTLVCLHNOEVTAEEDGTQRCCKSPCARVCYGLGMQYIKANSKFIGI 359  
 DB 300 PNYLSTDVSGCTLVCLHNOEVTAEEDGTQRCCKSPCARVCYGLGMQYIKANSKFIGI 359  
 QY 301 PNYLSTEVGSCCTLVCPNNQOEVTAEEDGTQRCCKSPCARVCYGLGMQYIKANSKFIGI 360  
 DB 301 PNYLSTEVGSCCTLVCPNNQOEVTAEEDGTQRCCKSPCARVCYGLGMQYIKANSKFIGI 360  
 QY 360 TELPACCKKIFGSLAPLPSFGDDPASNTAPLOPQVETLEETITGLYISAMPDSL 419  
 DB 360 TELPACCKKIFGSLAPLPSFGDDPASNTAPLOPQVETLEETITGLYISAMPDSL 419  
 QY 361 NQBEFGCKKIFGSLAPLPSFGDDPASNTAPLOPQVETLEETITGLYISAMPDSL 420  
 DB 361 NQBEFGCKKIFGSLAPLPSFGDDPASNTAPLOPQVETLEETITGLYISAMPDSL 420  
 QY 420 PDLSPONTLOVIRIRLHNGAYSLLTGLGISWLGRLSRELSGALIHNNHLCFVHT 479  
 DB 420 PDLSPONTLOVIRIRLHNGAYSLLTGLGISWLGRLSRELSGALIHNNHLCFVHT 479  
 QY 421 RDLSPONTLOVIRIRLHNGAYSLLTGLGISWLGRLSRELSGALIHNNHLCFVHT 480  
 DB 421 RDLSPONTLOVIRIRLHNGAYSLLTGLGISWLGRLSRELSGALIHNNHLCFVHT 480  
 QY 480 VPMDQCFERNPHQALHNRPEDE-CVGBGLACHQCARHCHGPGPTQCVNCSQTLRQ 538  
 DB 480 VPMDQCFERNPHQALHNRPEDE-CVGBGLACHQCARHCHGPGPTQCVNCSQTLRQ 538  
 QY 481 VPMDQCFERNPHQALHNRPEDE-CVGBGLACHQCARHCHGPGPTQCVNCSQTLRQ 540  
 DB 481 VPMDQCFERNPHQALHNRPEDE-CVGBGLACHQCARHCHGPGPTQCVNCSQTLRQ 540  
 QY 539 ECVEECRVLQGLPREYVNAHCLPCHECOPONGSVTCFPEADQVACAHYDPPFCVA 598  
 DB 539 ECVEECRVLQGLPREYVNAHCLPCHECOPONGSVTCFPEADQVACAHYDPPFCVA 598  
 QY 541 ECVEECRVLQGLPREYVNAHCLPCHECOPONGSVTCFPEADQVACAHYDPPFCVA 600  
 DB 541 ECVEECRVLQGLPREYVNAHCLPCHECOPONGSVTCFPEADQVACAHYDPPFCVA 600  
 QY 599 RCPBGVCPDLSTYPMIMKFPDEEGACCPINCHSCVDLDDKCPAOPRASPLTSVSAV 658  
 DB 599 RCPBGVCPDLSTYPMIMKFPDEEGACCPINCHSCVDLDDKCPAOPRASPLTSVSAV 658  
 QY 601 RCPBGVCPDLSTYPMIMKFPDEEGACCPINCHSCVDLDDKCPAOPRASPLTSVSAV 660  
 DB 601 RCPBGVCPDLSTYPMIMKFPDEEGACCPINCHSCVDLDDKCPAOPRASPLTSVSAV 660  
 QY 659 VGLLVVVLGVGFIILKRRQKIRKTYMRLLQETELVEPLTPSGAMPQAOMRIKKEF 718  
 DB 659 VGLLVVVLGVGFIILKRRQKIRKTYMRLLQETELVEPLTPSGAMPQAOMRIKKEF 718  
 QY 661 VGVLLFLVVLVVGFIILKRRQKIRKTYMRLLQETELVEPLTPSGAMPQAOMRIKKEF 720  
 DB 661 VGVLLFLVVLVVGFIILKRRQKIRKTYMRLLQETELVEPLTPSGAMPQAOMRIKKEF 720  
 QY 719 ANFVSWLKVPRKYSAS-HLETYKGIWIDGENVKIPVAIKVRENTSPKANKEILDEA 777  
 DB 719 ANFVSWLKVPRKYSAS-HLETYKGIWIDGENVKIPVAIKVRENTSPKANKEILDEA 777

DB 721 E-----LRKYKVGSAFGTIVKGIWIDGENVKIPVAIKVRENTSPKANKEILDEA 773  
 QY 778 YMAGVSPYVSRLLIGLCTISTVQVLTQMPYGLLDHRENRRLSGQDLNMCQIAK 837  
 DB 778 YMAGVSPYVSRLLIGLCTISTVQVLTQMPYGLLDHRENRRLSGQDLNMCQIAK 837  
 QY 774 YMAGVSPYVSRLLIGLCTISTVQVLTQMPYGLLDHRENRRLSGQDLNMCQIAK 833  
 DB 774 YMAGVSPYVSRLLIGLCTISTVQVLTQMPYGLLDHRENRRLSGQDLNMCQIAK 833  
 QY 838 GMSYEDVRLVHRDLAANVLYKSPNHYKIDFGIARLLDIDETEHADGKVPKMMAL 897  
 DB 838 GMSYEDVRLVHRDLAANVLYKSPNHYKIDFGIARLLDIDETEHADGKVPKMMAL 897  
 QY 834 GMSYEDVRLVHRDLAANVLYKSPNHYKIDFGIARLLDIDETEHADGKVPKMMAL 893  
 DB 834 GMSYEDVRLVHRDLAANVLYKSPNHYKIDFGIARLLDIDETEHADGKVPKMMAL 893  
 QY 898 ESIILRRFTHSDVMSYVTVWELMTFCAPKYDGIAPAEIPDLLEKGERLPOPICTIDV 957  
 DB 898 ESIILRRFTHSDVMSYVTVWELMTFCAPKYDGIAPAEIPDLLEKGERLPOPICTIDV 957  
 QY 894 ESIILRRFTHSDVMSYVTVWELMTFCAPKYDGIAPAEIPDLLEKGERLPOPICTIDV 953  
 DB 894 ESIILRRFTHSDVMSYVTVWELMTFCAPKYDGIAPAEIPDLLEKGERLPOPICTIDV 953  
 QY 958 YMIWYKCMWIDSECPREFRELVEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLED 1017  
 DB 958 YMIWYKCMWIDSECPREFRELVEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLED 1017  
 QY 954 YMIWYKCMWIDSECPREFRELVEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLED 1013  
 DB 954 YMIWYKCMWIDSECPREFRELVEFSRMAADPQRFVVIQNEIDLGPASPLDSTFYRSLLED 1013  
 QY 1018 DDMGDLVDAEYLVPOQGFPCDDPAPAGGVVHHRSSSTRSGGDLTLGLEPSEEEAP 1077  
 DB 1018 DDMGDLVDAEYLVPOQGFPCDDPAPAGGVVHHRSSSTRSGGDLTLGLEPSEEEAP 1077  
 QY 1014 DDMGDLVDAEYLVPOQGFPCDDPAPAGGVVHHRSSSTRSGGDLTLGLEPSEEEAP 1073  
 DB 1014 DDMGDLVDAEYLVPOQGFPCDDPAPAGGVVHHRSSSTRSGGDLTLGLEPSEEEAP 1073  
 QY 1078 RSLPAPSEGAGSDVFDGDLGMAKGLQSLPTHDSPLQRYSEDFTVPLPESTDGYVAPL 1137  
 DB 1078 RSLPAPSEGAGSDVFDGDLGMAKGLQSLPTHDSPLQRYSEDFTVPLPESTDGYVAPL 1137  
 QY 1074 RSLPAPSEGAGSDVFDGDLGMAKGLQSLPTHDSPLQRYSEDFTVPLPESTDGYVAPL 1133  
 DB 1074 RSLPAPSEGAGSDVFDGDLGMAKGLQSLPTHDSPLQRYSEDFTVPLPESTDGYVAPL 1133  
 QY 1138 TCSPOPEYVNOQDVARPOPPREGLPAPAPAGATLEPAKTLSPCKNVVVDVFAFGAV 1197  
 DB 1138 TCSPOPEYVNOQDVARPOPPREGLPAPAPAGATLEPAKTLSPCKNVVVDVFAFGAV 1197  
 QY 1134 ACSPOPEYVNOQDVARPOPPREGLPAPAPAGATLEPAKTLSPCKNVVVDVFAFGAV 1193  
 DB 1134 ACSPOPEYVNOQDVARPOPPREGLPAPAPAGATLEPAKTLSPCKNVVVDVFAFGAV 1193  
 QY 1198 ENPEYLTPOQGAAPPHPPAFSPAFDMLYVWDQDPPRGAAPSTFTKPTAENEEYGL 1257  
 DB 1198 ENPEYLTPOQGAAPPHPPAFSPAFDMLYVWDQDPPRGAAPSTFTKPTAENEEYGL 1257  
 QY 1194 ENPEYLTPOQGAAPPHPPAFSPAFDMLYVWDQDPPRGAAPSTFTKPTAENEEYGL 1253  
 DB 1194 ENPEYLTPOQGAAPPHPPAFSPAFDMLYVWDQDPPRGAAPSTFTKPTAENEEYGL 1253  
 QY 1258 DVPV 1261  
 DB 1258 DVPV 1261  
 QY 1254 DVPV 1257  
 DB 1254 DVPV 1257

RESULT 3  
 ERB2\_MESAU STANDARD; PRT: 1254 AA.  
 ID ERB2\_MESAU  
 AC Q60553;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DE 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).  
 GN ERB2 OR NEU.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 OX NCBI\_Taxid=10036;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Nerve;  
 RX MEDLINE=94193007; PubMed=7908275;  
 RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,  
 RT Yamazaki Y., Ishikawa T.;  
 RT Cloning and activation of the Syrian hamster neu proto-oncogene.";  
 RL Gene 140:251-255(1994).  
 CC - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.  
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A  
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).  
 CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERB2 RECEPTORS  
 CC (POTENTIAL).  
 CC - SUBCELLULAR LOCATION: Type I membrane protein.  
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE



CC RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; D16295; BAA03801.1; -  
 DR HSSP; P11362; IEGK.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_Pkinase.  
 DR InterPro; IPR004019; YLP\_motif.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR Pfam; PF02757; YLP; 2.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Proto-oncogene; Disease mutation.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 1 21 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT DOMAIN 22 1254 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 653 675 POTENTIAL.  
 FT DOMAIN 676 1254 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 158 368 CYS-RICH.  
 FT DOMAIN 472 644 CYS-RICH.  
 FT DOMAIN 720 987 PROTEIN KINASE.  
 FT NP\_BIND 726 734 ATP (BY SIMILARITY).  
 FT BINDING 753 753 ATP (BY SIMILARITY).  
 FT ACT\_SITE 845 845 BY SIMILARITY.  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 236 244 BY SIMILARITY.  
 FT DISULFID 240 252 BY SIMILARITY.  
 FT DISULFID 255 264 BY SIMILARITY.  
 FT DISULFID 268 295 BY SIMILARITY.  
 FT DISULFID 299 311 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 511 520 BY SIMILARITY.  
 FT DISULFID 515 528 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 544 560 BY SIMILARITY.  
 FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 587 584 BY SIMILARITY.  
 FT DISULFID 596 623 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1247 1247 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 125 125 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARIANT 658 658 V -> E (IN ONCOGENIC NEU).  
 FT VARIANT 659 659 V -> E (IN ONCOGENIC NEU).  
 SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2BB1 CRC64;

Query Match 85.1%; Score 5834.5; DB 1; Length 1254;  
 Best Local Similarity 85.3%; Pred. No. 2-2e-305;  
 Matches 1077; Conservative 61; Mismatches 115; Indels 9; Gaps 3;  
 1 MELALACRWGLLALIPGAASTQVCTGDMRLRPAPBETHLDMLRHLYQCCQVQGNL 60  
 1 MELAMCGGLLALLSPASQVCTGDMRLRPAPBETHLDMLRHLYQCCQVQGNL 60  
 61 ELTYLPTNASSLPLODIOEVQGVYLIANQVQVPLQRLIRYRGTLQLEEDNALAVLNG 120  
 61 ELTYLPANATLSFLDIOEVQGVYLIANQVQVPLQRLIRYRGTLQLEEDNALAVLNG 120  
 121 DELNNTTPVTGASPGGLRELQRLSTLEILKGVLLQRPDQCYODTIIMKDIFFKNQOLA 180  
 121 DELDNTTATGTTPGELRELQRLSTLEILKGVLLQRPDQCYODTIIMKDIFFKNQOLA 180  
 181 LTLIDTNRSRACHPSPCMKGSRCWGESSEDCQSLTRTVAGGACRCCKGPLEPTDCHEQC 240  
 181 PVDIDTNRSRACHPSPCMKGSRCWGESSEDCQSLTRTVAGGACRCCKGPLEPTDCHEQC 240  
 241 AAGCTGPKKSDCLALCFHNSGICELHCPALVTNTDTPESMPNPEGRTFGASCVTCP 300  
 241 AAGCTGPKKSDCLALCFHNSGICELHCPALVTNTDTPESMPNPEGRTFGASCVTCP 300  
 301 YNYLSTDVGSCTLVCPHNOEYTAEDGTQRCCKSPCARVCYGLMAYIKANSKFIGIT 360  
 301 YNYLSTDVGSCTLVCPHNOEYTAEDGTQRCCKSPCARVCYGLMAYIKANSKFIGIT 360  
 361 ELFPAGCKKIFSSLAFLPESFGDPASTAPLQPOLOVFEETLITGLYISAMPDILP 420  
 361 ELFPAGCKKIFSSLAFLPESFGDPASTAPLQPOLOVFEETLITGLYISAMPDILP 420  
 421 IDEFAGCKKIFSSLAFLPESFGDPASTAPLQPOLOVFEETLITGLYISAMPDILP 420  
 421 IDEFAGCKKIFSSLAFLPESFGDPASTAPLQPOLOVFEETLITGLYISAMPDILP 420  
 421 DLSVQNLQVIRIGRIHNGAYSLTQGLISLGLRSRLRELSGALVHNNHLCFVHTV 480  
 421 DLSVQNLQVIRIGRIHNGAYSLTQGLISLGLRSRLRELSGALVHNNHLCFVHTV 480  
 481 PWDQLFRNHQALLHTANRPEDECVGEGLAGLQLCARHGWGPGPTQCVNCSQFLRGEC 540  
 481 PWDQLFRNHQALLHTANRPEDECVGEGLAGLQLCARHGWGPGPTQCVNCSQFLRGEC 540  
 481 PWDQLFRNHQALLHTANRPEDECVGEGLAGLQLCARHGWGPGPTQCVNCSQFLRGEC 540  
 481 PWDQLFRNHQALLHTANRPEDECVGEGLAGLQLCARHGWGPGPTQCVNCSQFLRGEC 540  
 541 VEECVLQGLPREVYNAHCLPCHPECPONSVCDFEADQCAACHKCPKPCVAC 600  
 541 VEECVLQGLPREVYNAHCLPCHPECPONSVCDFEADQCAACHKCPKPCVAC 600  
 541 VEECVLQGLPREVYNAHCLPCHPECPONSVCDFEADQCAACHKCPKPCVAC 600  
 541 VEECVLQGLPREVYNAHCLPCHPECPONSVCDFEADQCAACHKCPKPCVAC 600  
 601 PEGVAPDLSTYMPIMKPEDEBEGACQPCPINTCHSCVDLDKGPACQBPASPLTSIYAVG 660  
 601 PEGVAPDLSTYMPIMKPEDEBEGACQPCPINTCHSCVDLDKGPACQBPASPLTSIYAVG 660  
 601 PEGVAPDLSTYMPIMKPEDEBEGACQPCPINTCHSCVDLDKGPACQBPASPLTSIYAVG 660  
 601 PEGVAPDLSTYMPIMKPEDEBEGACQPCPINTCHSCVDLDKGPACQBPASPLTSIYAVG 660  
 661 ILVVVGVVFGIILKRRQOKIRKTYMRLLOETELVBPPLPSGAMPNOAOKRIKERN 720  
 661 ILVVVGVVFGIILKRRQOKIRKTYMRLLOETELVBPPLPSGAMPNOAOKRIKERN 720  
 661 ILVVVGVVFGIILKRRQOKIRKTYMRLLOETELVBPPLPSGAMPNOAOKRIKERN 720  
 661 ILVVVGVVFGIILKRRQOKIRKTYMRLLOETELVBPPLPSGAMPNOAOKRIKERN 720  
 721 FTVSFMLRPPKYSAS-HLETYKGIWIPDEGVKIPVAIKVLRNTSPKANEILDEAYV 779  
 721 FTVSFMLRPPKYSAS-HLETYKGIWIPDEGVKIPVAIKVLRNTSPKANEILDEAYV 779  
 721 FTVSFMLRPPKYSAS-HLETYKGIWIPDEGVKIPVAIKVLRNTSPKANEILDEAYV 779  
 721 FTVSFMLRPPKYSAS-HLETYKGIWIPDEGVKIPVAIKVLRNTSPKANEILDEAYV 779  
 780 MAGVSPYVSRLLGICLTSTVOLVTLMPYGGCLDHYVBNRGLCSODLLNMCQIANG 839  
 780 MAGVSPYVSRLLGICLTSTVOLVTLMPYGGCLDHYVBNRGLCSODLLNMCQIANG 839  
 780 MAGVSPYVSRLLGICLTSTVOLVTLMPYGGCLDHYVBNRGLCSODLLNMCQIANG 839  
 780 MAGVSPYVSRLLGICLTSTVOLVTLMPYGGCLDHYVBNRGLCSODLLNMCQIANG 839  
 840 SYLEQVRLVHRDLARNLVYVSPNFKITTDGLARLLIDEFYVADGKPIKMALES 899  
 840 SYLEQVRLVHRDLARNLVYVSPNFKITTDGLARLLIDEFYVADGKPIKMALES 899  
 840 SYLEQVRLVHRDLARNLVYVSPNFKITTDGLARLLIDEFYVADGKPIKMALES 899  
 840 SYLEQVRLVHRDLARNLVYVSPNFKITTDGLARLLIDEFYVADGKPIKMALES 899  
 894 ILRRFTHQSDVWSGYVTWELMTGAPYDGIIPAREIPDLLEKGERLPQPICTIDYVM 953  
 894 ILRRFTHQSDVWSGYVTWELMTGAPYDGIIPAREIPDLLEKGERLPQPICTIDYVM 953  
 894 ILRRFTHQSDVWSGYVTWELMTGAPYDGIIPAREIPDLLEKGERLPQPICTIDYVM 953  
 894 ILRRFTHQSDVWSGYVTWELMTGAPYDGIIPAREIPDLLEKGERLPQPICTIDYVM 953  
 954 IMVKCMIDSECRPRELVESEFARMADPQRFVIONEDIGPASPLDSTYRSLTEDD 1013  
 954 IMVKCMIDSECRPRELVESEFARMADPQRFVIONEDIGPASPLDSTYRSLTEDD 1013  
 954 IMVKCMIDSECRPRELVESEFARMADPQRFVIONEDIGPASPLDSTYRSLTEDD 1013  
 954 IMVKCMIDSECRPRELVESEFARMADPQRFVIONEDIGPASPLDSTYRSLTEDD 1013

QY 1020 MGDLVDAEYLVPOGGFFCPDAPGAGMVRHRRSSSTRSGGDLTGLPSSSEAPRS 1079  
 DB 1014 MGDLVDAEYLVPOGGFFPDAPAGSTARRRRSSSTRGGDLTGLMPSSEPPRS 1073  
 QY 1080 PLAPSEGAGSDVDFDGLGMAKAGLQSLPTDPSPLQRYSDPVPPLPSETDGVAPLTC 1139  
 DB 1074 PLAPSEGAGSDVDFEELMGATKQGSISPRDLSPLOKRYSDPPLPSETDGVAPLAC 1133  
 QY 1140 SPOPEYVNOVDVRFPPSPRPGPLPAPAPACATLTERAKTSLSPGXGVYKGVFAFGAVEN 1199  
 DB 1134 SPOPEYVNOVDVRFPPSPRPGPLPAPAPACATLTERAKTSLSPGXGVYKGVFAFGAVEN 1193  
 QY 1200 PEVLTPOGGAPOPPHPAPSPAPDNLYYWDQDPPRCAPSTFKGTPTAENPEYLGIDV 1259  
 DB 1194 PEVLTPOGGAPOPPHPAPSPAPDNLYYWDQDPPRCAPSTFKGTPTAENPEYLGIDV 1252  
 QY 1260 PV 1261  
 DB 1253 PV 1254  
 RESULT 4  
 EGRF\_HUMAN STANDARD. PRT; 1210 AA.  
 ID EGRF\_HUMAN  
 AC P00533; P06268; Q14225; Q9CMD7; Q9UMD8; Q9UMG5; Q92795; O00732;  
 AC O00688; Q9B2S2; Q9H2C9; Q9G2X1; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Receptor  
 protein-tyrosine kinase ErbB-1).  
 GN EGRF OR ERBB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=84219729; PubMed=6328312;  
 RA Ullrich A., Conness L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
 Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,  
 Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;  
 RA "Human epidermal growth factor receptor cDNA sequence and aberrant  
 RT expression of the amplified gene in A431 epidermoid carcinoma cells";  
 RL Nature 309:418-425(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=95382957; PubMed=7654368;  
 RA Ilekis J.V., Stark B.C., Scoccia B.;  
 RT "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta";  
 RL Mol. Reprod. Dev. 41:149-156(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=97076686; PubMed=8918811;  
 RA Reiter J.L., Mahle N.J.;  
 RT "A 1.6 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor";  
 RL Nucleic Acids Res. 24:4050-4056(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RX MEDLINE=97256547; PubMed=9103388;  
 RA Ilekis J.V., Gariti J., Niederberger C., Scoccia B.;  
 RT "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGR) in ovarian cancer";  
 RL Gynecol. Oncol. 65:36-41(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RX TISSUE=Placenta;  
 RX MEDLINE=21100872; PubMed=11161793;

RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 RA Scheel Sinclair C., Pearseall R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balaubramaniam S., Crossley T.D., Magnuson T.R., James C.D.,  
 RA Mahle N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGRF transcripts encoding truncated receptor  
 RT isoforms";  
 RL Genomics 71:1-20(2001).  
 RN [6]  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Scheel C.M.,  
 RA Lampland A.L., Balaubramaniam S., Crossley T.O., Magnuson T.R.,  
 RA Mahle N.J.;  
 RT "Human and mouse alternative EGRF transcripts encoding only the  
 RT extracellular domain of the receptor";  
 RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=64196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krutiger W., Stolarsky L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;  
 RT "Expression cloning of human EGF receptor complementary DNA: gene  
 RT amplification and three related messenger RNA products in A431  
 RT cells";  
 RL Science 224:843-848(1984).  
 RN [8]  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.;  
 RT "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAs overproduced in A431 carcinoma cells";  
 RL Nature 309:806-810(1984).  
 RN [9]  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Symmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,  
 RA O'Malley B.W.;  
 RT "Isolation of an evolutionarily conserved epidermal growth factor  
 RT receptor cDNA from human A431 carcinoma cells";  
 RL Biochem. Biophys. Res. Commun. 124:1125-1132(1984).  
 RN [10]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kitchington D., Ullrich A.,  
 RA Waterfield M.D.;  
 RT "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription";  
 RL Oncogene Res. 1:375-386(1987).  
 RN [11]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RT "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RT factor receptor proto-oncogene RNA synthesis";  
 RL J. Biol. Chem. 266:1746-1753(1991).  
 RN [12]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2991899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RT "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN [13]  
 RP SEQUENCE OF 540.  
 RA Kohda D.;  
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RN [14]  
 RP RECEPTOR ACTIVITY  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Mroczkowski B., Mosig G., Cohen S.;  
 RT "ATP-stimulated interaction between epidermal growth factor receptor

RT and supercoiled DNA." ;  
 RL Nature 309:270-273(1984).  
 (15)  
 RP PHOSPHORYLATION  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L.; Lax I.; Kris R.; Dombalagian M.; Honegger A.M.;  
 RA Howk R.; Glyvol D.; Ullrich A.; Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)  
 RT receptor and HER2/neu are located in their carboxyl-terminal tails.  
 RT Identification of a novel site in EGF receptor." ;  
 RL J. Biol. Chem. 264:10667-10671(1989).  
 (16)  
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RX MEDLINE=96398132; PubMed=8962717;  
 RA Smith K.D.; Davies M.J.; Bailey D.; Renouf D.V.; Hounsell E.F.;  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 RT the epidermal growth factor receptor expressed in Chinese hamster  
 RT ovary fibroblasts." ;  
 RL Growth Factors 13:121-132(1996).  
 (17)  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; PubMed=10731668;  
 RA Sato C.; Kim J.-H.; Abe Y.; Sato K.; Yokoyama S.; Kohda D.;  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor  
 RT receptor." ;  
 RL J. Biochem. 127:65-72(2000).  
 (18)  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=9825196; PubMed=9556602;  
 RA Abe Y.; Odaoka M.; Inagaki F.; Lax I.; Schlessinger J.; Kohda D.;  
 RT "Disulfide bond structure of human epidermal growth factor receptor." ;  
 RL J. Biol. Chem. 273:11150-11157(1998).  
 (19)  
 RP REVIEW.  
 RX MEDLINE=87297456; PubMed=3039909;  
 RA Carpenter G.;  
 RT "Receptors for epidermal growth factor and other polypeptide  
 RT mitogens." ;  
 RL Annu. Rev. Biochem. 56:881-914(1987).  
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF  
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 CC EGF-like growth factor, Gp10 and vaccinia virus growth factor. It  
 CC is involved in the control of cell growth and differentiation.  
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 CC secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/  
 CC truncated isoform/TBGR, 3/p110 and 4; are produced by  
 CC alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 CC expressed in ovarian cancers.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@sdb.ch](mailto:license@sdb.ch)).  
 CC -----  
 DR EMBL; X00588; CAA25240.1; -;  
 DR EMBL; U95089; AAB53063.1; -;  
 DR EMBL; U48722; AAC50802.1; -;

DR EMBL; U48723; AAC50804.1; -;  
 DR EMBL; U48724; AAC50796.1; -;  
 DR EMBL; U48725; AAC50797.1; -;  
 DR EMBL; U48726; AAC50798.1; -;

## Query Match

44.7% Score 3063; DB 1; Length 1210;  
 Best Local Similarity 48.7%; Pred. No. 7e-157;

Matches 620; Conservative 178; Mismatches 360; Indels 114; Gaps 24;

11 LLLALLPQGA--STQVCTGDMKRLRLASPTHTLDMRLHLYQCGVQVQNELTYLPN 68  
 14 LLAALCPASRLAEKRVQGSNKLTLQGTEDHFLSLQRMFNNEVVLGNLEITYVGRN 73  
 69 ASLSFLQDIOEYGVYLAHNVROVPLORIRIVYGTOLFEDNYVLAVIDNDPLNNTP 128  
 74 YDLSLTKIQEAVGVYLAHNVRIPLNQLIRGNMYNSYLAVALSYND----- 126  
 129 VTGASPGGLRELQSLTEIILKGVLIQRNPQLCYDTILMKDIFKNNQALALIDITNR 188  
 127 --ANKTLKELPMENLOEILHGAVFSSNPALCVESIQWRDIVSDFLSNMSMDFNH 183  
 189 SRACHPGSPMCKGSRCWGESSSDQSLRTVCAGCA--RCKGPLETDCHEQCAAGCTP 247  
 184 LSCCKCPSCPNCSWAGBNCKLTKIICAQCCSRCKGKSPSCCHNOCAAGCTP 243  
 248 KHSDDLACIHNHSGICELHCPALVTYNTDFFESMPNDEGRYTFGASCVTACPYNYLST 307  
 244 RESDLVCRKFRDEATCKDTCPLMLVNPPTYQMDVNEGKYSFGATCVKCCPRVYVTD 303  
 308 VQSCITVCPLNQETADTGTCRCEKSPCARVYGLGMQYIKANSFGLTTELE--FAG 366  
 304 HGSVCVRACGADSYEN--EEDGVKCKCKGPRKVCNNGIGEFK--DSLSINATNKHFN 361  
 367 CKTIKFSALFPESHGDPASNTAPLQEOQLVETLEETLEIGYLYIASMPDPLPLSFQ 426  
 362 CTSISGDHLILPVAFRGDSFHTPLDQELDLKTYEITGLFLLIQMPENRDLDAFE 421  
 427 NLQVIRGILHNHGVSLTLQGLISWGLRLRLRELISGLALIHNTLFCYHTVPMQDLF 486  
 422 NLEIRGRTKHQGFSLAVSILNITSLGRSLKEISDDVILISGNKNCVANTIMWKLF 481  
 487 RNPQALLHTANRPDECVGSLACHOLCAHGCHMGPPTQCVNCSQSLROECVEEGRV 546  
 482 GTSQKTKIISNGENSKATGOVCHALCSPEGCGPRPDCSRNYSRGRBCVDKNL 541  
 547 LQGLPREYVNAHCLPCHPEQPONGSVTCGPPEADQCVACAHYKDPFCVACRPSGVK 606  
 542 LEGEPRFEVNSECTIOCHPECLPQAMNITCTGRGPDNICOAHYIDPHCVKTCGAVWG 601  
 607 DLSYPIWKFPDEGACQPCPINCTHSCVDDDDGCPAEQASPLTSIVSAVG--ILL 663  
 602 ENNTL-VKTVADAGVCHLCPNCTYGTGGLGCGCPNNGKIP--STATMVALLLL 658  
 664 VVVLGVVFGILIKRQCKIRKYTRRLLOETELVEPLTPSGAMPNOAMRLKEFNNTV 723  
 659 VVALGIG---LPMRRRHIVKRTLRLLQHELEVEPLTPSEARNQALRLIKE----- 709  
 724 SFMLRVPVNSHSETYTKGVIWPDGENVKIPVAIKLRENTSPRANKELIDEAYVAVG 783  
 710 TEFKIKIGLAGAGTYKGLMPEGEVYKIPVAIKELRENTSPRANKELIDEAYVAVS 769  
 784 GSPVSRLLGICLSTVQVLTQMPYGLDHDVENGRGLSGDOLLMQMOIAGMSYLE 843  
 770 DNPVYCHLGLISTVQVLTQMPYGLDHDVENGRGLSGDOLLMQMOIAGMSYLE 829  
 844 DVRLVHDLAARVLYKSPNHVKTITDGLARLLIDETEHADGKVPKIMMALESILRR 903  
 830 DRLVHDLAARVLYKSPNHVKTITDGLARLLIDETEHADGKVPKIMMALESILRR 889  
 904 RFTQSDVMSYGVVLMETFGAKPYDGIAREIPDLLENGERLPOPEICTIDVYIMVK 963  
 890 IYTHQSDVMSYGVVLMETFGAKPYDGIAREIPDLLENGERLPOPEICTIDVYIMVK 949

QY 964 CWMIDSECRPRELVESEFMRMDPQRFVIO-NEDLGPASPLDSTFYRLIEDDDMGD 1022  
 DB 950 CWMIDADSRKPELLEFSEKMDPQRFVIOGDERMLPSTDSNFTALMDDEMDMD 1009  
 QY 1023 LVDAEYLVEQGFCCFPDPAFGAGVHHRHRSSTRSGGDLTLLGSEEEAPRPLA 1082  
 DB 1010 VVDADEYLVEQGF-----SSPSTRTPL 1035  
 QY 1083 PSEGAGSDVFDGLGMAAGLQSLPHDPSLQRYSEDPVLPSET--DGVAVLTS 1140  
 DB 1036 SLSLSTSN--NSTVACIDRNGLOSCPKEDSFQRYSSDPTGALTEDSIDTFL----- 1087  
 QY 1141 POPEYVQPDVRFQPSFREGPLPAARPAATLERAKTLSPGKGVKGVFAFGAVENP 1200  
 DB 1088 FVEEYINQ-SVPRKPRAGSVQNPVYHNPQNP-----APSRDPHYQD--PRSTAVGNP 1136  
 QY 1201 EYL-TPQGAAPCPHPPPAFSPAFDNLVYWDQ-----DP-----PERGAPPSFF 1243  
 DB 1137 EYINTVQ-----PTCVNSTFDSFAMWAKGSHQISLNDPDYQDFFPKAKENGIF 1187  
 QY 1244 KGTPTAENPEYL 1255  
 DB 1188 KGS-TAENAEYL 1198

RESULT 5  
 EGFR\_MOUSE  
 ID EGFR\_MOUSE STANDARD; PRT; 1210 AA.  
 AC 001279;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112).  
 GN EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C; TISSUE=Liver;  
 RA MEDLINE=93026370; Pubmed=1408137;  
 RA Avivi A., Skorecki K., Yavon A., Givol D.;  
 RT "Promoter region of the murine fibroblast growth factor receptor 2  
 RT (bek/KGFR) gene.";  
 RL Oncogene 7:1957-1962 (1992).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C, and CD-1; TISSUE=Liver, and Decidua;  
 RX MEDLINE=93126380; Pubmed=7678348;  
 RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;  
 RT "Expression of the epidermal growth factor receptor gene is regulated  
 RT in mouse blastocysts during delayed implantation.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59 (1993).  
 RN (3)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BALB/C; TISSUE=Liver;  
 RA Hibbs M.L.;  
 RT Submitted (APR-1994) to the EMBL/GenBank/DBJ databases.  
 RN (4)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B6/C3; TISSUE=Liver;  
 RX MEDLINE=94170986; Pubmed=6125255;  
 RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Earp H.S.,  
 RA Jenkins N.A., Lee D.C.;  
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF  
 RT receptor tyrosine kinase.";  
 RL Genes Dev. 8:339-413 (1994).  
 RN (5)  
 RP SEQUENCE OF 1-714 FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=91232866; Pubmed=2030916;  
 RA Avivi A., Lax I., Ullrich A., Schlessinger J., Givol D., Morse B.;

RT "Comparison of EGF receptor sequences as a guide to study the ligand  
 RT binding site.";  
 RT Oncogene 6:673-676 (1991).  
 RN (6)  
 RN SEQUENCE OF 969-1117 FROM N.A.  
 RP STRAIN=C3H;  
 RA Bissinger D.P., Serrero G.;  
 RT Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,  
 CC AND ALSO OF TGF-ALPHA, AMPHIRGULIN, HEPARIN-BINDING EGF, GP30 AND  
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; X78987; CAA5587.1; -  
 DR EMBL; U03425; AAA17899.1; -  
 DR EMBL; X59698; CAA42219.1; -  
 DR EMBL; L06864; AAA53029.1; -  
 DR EMBL; Z12608; CAA78249.1; -  
 DR HSSP; P11362; 1FGK.  
 DR MGD; MGI:95294; Egfr.  
 DR InterPro; IPR00494; EGFR\_L domain.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR002174; Purin-like.  
 DR InterPro; IPR001245; Tyr\_Pkinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF01030; Recep\_L domain; 2.  
 DR ProDom; PD000601; Euk\_Pkinase; 1.  
 DR SMART; SMO0261; Fu; 3.  
 DR SMART; SMO0219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 DR PROSITE; PS00011; PROTEIN KINASE DOM; 1.  
 KW Transmembrane; Glycoprotein; Receptor; Signal; Transferrase;  
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.  
 FT SIGNAL 1 24  
 FT CHAIN 25 1210  
 FT DOMAIN 25 647  
 FT TRANSMEM 648 670  
 FT DOMAIN 671 1210  
 FT REPEAT 75 300  
 FT REPEAT 390 600  
 FT DOMAIN 1028 1071  
 FT DOMAIN 714 981  
 FT NP\_BIND 720 728  
 FT BINDING 747 747  
 FT ACT\_SITE 839 839  
 FT DISULFID 190 199  
 FT DISULFID 194 207  
 FT DISULFID 215 223  
 FT DISULFID 219 231  
 FT DISULFID 232 240  
 FT DISULFID 236 248  
 FT DISULFID 251 260  
 FT DISULFID 264 291  
 FT DISULFID 295 307  
 FT DISULFID 311 326  
 FT DISULFID 329 333  
 FT DISULFID 333 333

```

FT DISULFID 506 515 BY SIMILARITY.
FT DISULFID 510 523 BY SIMILARITY.
FT DISULFID 526 535 BY SIMILARITY.
FT DISULFID 539 555 BY SIMILARITY.
FT DISULFID 558 571 BY SIMILARITY.
FT DISULFID 562 579 BY SIMILARITY.
FT DISULFID 582 591 BY SIMILARITY.
FT DISULFID 595 617 BY SIMILARITY.
FT DISULFID 620 628 BY SIMILARITY.
FT DISULFID 624 636 BY SIMILARITY.
FT MOD_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).
FT MOD_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1110 1110 PHOSPHORYLATION (BY SIMILARITY).
FT MOD_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT MOD_RES 1197 1197 PHOSPHORYLATION (AUTO-, MAJOR SITE) (BY SIMILARITY).
FT CARBOHYD 128 128 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 175 175 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 196 196 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 413 413 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 528 528 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 603 603 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 623 623 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CONFLICT 19 19 C -> S (IN REF. 2).
FT CONFLICT 539 539 C -> W (IN REF. 5).
FT CONFLICT 991 991 L -> F (IN REF. 4).
FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).
SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DF222F5 CRC64;

```

Query Match 44.3%; Score 3037; DB 1; Length 1210;  
 Best Local Similarity 48.6%; Pred. No. 1,7e-155;  
 Matches 622; Conservative 172; Mismatches 367; Indels 118; Gaps 26;

```

QY 11 LLLALLPPGAA--STOYCTGDMKRLPASPETHLDMWRHLYOCGVVQGNLETTYPTN 68
DB 14 LLLALLPPGAA--STOYCTGDMKRLPASPETHLDMWRHLYOCGVVQGNLETTYPTN 73
QY 69 ASLSFLDIOEVQVYLIANQVROVPLQRLIRVQGLEEDNAYALVLDNGPPLNNTTP 128
DB 74 YLSFLKTIQEVQVYLIANQVROVPLQRLIRVQGLEEDNAYALVLDNGPPLNNTTP 124
QY 129 VTGSPGRLRELQRLITLILKGVLLQQRNPQLQYQDTIIMKDI-----FKNNQALATLI 164
DB 125 -YGINRTGLRELRLKRLQRLITLILKGVLLQQRNPQLQYQDTIIMKDI-----FKNNQALATLI 180
QY 185 DTNRSRACHPCSPKSGRCKGSESSDQSLRTVTCAGGA-RCKGSLPTDCHEQCAAG 243
DB 181 -QSHPSSCPCKDPCSPKSGRCKGSESSDQSLRTVTCAGGA-RCKGSLPTDCHEQCAAG 239
QY 244 CTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMNPREFRYTFGASCTYACPNY 303
DB 240 CTGPKHSDCLACHFNHSGICELHCPALVYNTDTFESMNPREFRYTFGASCTYACPNY 299
QY 304 LSTDVSGCTLVCPPLHNOEYTAEDGTQRCCKSKCAVCGYGLQMOYKASKSTIGITELE 363
DB 300 VSTDVSGCTLVCPPLHNOEYTAEDGTQRCCKSKCAVCGYGLQMOYKASKSTIGITELE 357
QY 364 -FAGCKKIFGSLALPESFDGDPKSNAPLQPELOQVFTLEBTIGLYISAMPDSLPLD 422
DB 358 HFXYCTAISGLHLPLVAFKGDSTRTPLPDLPRELEILKTKVETIGLTLQAMPDWTDL 417
QY 423 SVFQNLQVIRGRILHNGAYSLTQGLISWGLRSLELSGLALIHNNTHLFCVTPW 482
DB 418 HAFENLEIRGRITQHGQFSLAVGUNTISGLRSLEISDGVIIISGNNTLCYANTIM 477
QY 483 DQFRNHQALLHTANPEDECEGELACIQLCARHCGWGPPTQVNCQSGILGECVE 542
DB 478 KKLFGTNGQTKIKMNNNAEKDCAVNVHCNPLSSSECCWGPPEPDDCVSCNVSREGREVE 537
QY 543 ECRVLQGLPREYVAVARHCLPCHPECQFONGSVTCFGEADQCAVCAHYKPPPCVARGCS 602

```

```

DB 538 KNVILGGEREFRENESECICQCHCECLPQAMNITCTGRGDNCCQCHAYIDGPRCTCPA 597
QY 603 GYKPDLSWYPIKWEDEBACOPCPINCHSCVDLDKQCPAEQASPLTSVSAVGL 662
DB 598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQCEVWPSGPKIPSIATGVGL 656
QY 663 LVYVGLVNGI-LIKRQOKIRKYTRRLLOETELVEPLTPSGAMNQOMRLKEPNF 721
DB 657 LFLV-VVAGIGLPMRRRIIVKRIITRLLORELEVEPLTPSGAMNQOMRLKEPNF 711
QY 722 TVSFMLRVKVSASHLETYYKGIWIPDGENVKIPIVAKYLRNTSPKANKELDEAYVA 781
DB 712 --TEFKIKVILGAGAGTVYKGIWIPDGENVKIPIVAKYLRNTSPKANKELDEAYVA 769
QY 782 GVSPFVSRLGLCLTSTVQVLTQMLPFGCLLDHREANGRLASQDILWMCQIAGMSY 841
DB 770 SYDNPFVCRLLGLCLTSTVQVLTQMLPFGCLLDHREANGRLASQDILWMCQIAGMSY 829
QY 842 LEDVRLVHRDLAARVLYKSPNHVKITDGLRLDIDETRYHADGKYPIMMALESIL 901
DB 830 LEDVRLVHRDLAARVLYKSPNHVKITDGLRLDIDETRYHADGKYPIMMALESIL 889
QY 902 RRFVTHQSDVMSYGVTVWELMTFGAKPYGIPARSLPDLLEKGERLPQPICTIDVYTM 961
DB 890 HRYTHQSDVMSYGVTVWELMTFGAKPYGIPARSLPDLLEKGERLPQPICTIDVYTM 949
QY 962 VKCMMDISCRPFRELVESEFSRMAADPQFVIO-NEDLGPASPLDSTFFYSLEDDM 1020
DB 950 VKCMMDISCRPFRELVESEFSRMAADPQFVIO-NEDLGPASPLDSTFFYSLEDDM 1009
QY 1021 GDLVDAEYLVPQGFCEPDPAPAGAGVYHHRSSSTRSGGDLTLGLEPSEEAASP 1080
DB 1010 EDVVDADDEVILIQGFEE-----NSPST-----SRTP 1035
QY 1081 LAPSEAGSDVDGDLGMAAGKGLQSLPHDSPLQRYBEDPVLPSPT--DGYAPRL 1138
DB 1036 LSSLSATSN---NSTVACINRNGSCRYKEDAFQRYSSDPGATTEINDIDAFV--- 1087
QY 1139 CSPOEYVQNPVPRPQPSRGPPLAARPAATLERAKTSLPGKGVVKKVAFAGAVE 1198
DB 1088 --PVPEYVQNPVPRPQPSRGPPLAARPAATLERAKTSLPGKGVVKKVAFAGAVE 1134
QY 1139 NPEYLT-TPGGAAPQHPPPASPAFNUITYDQ-----DP-----PERGAPS 1241
DB 1135 NPEYLT-TPGGAAPQHPPPASPAFNUITYDQ-----DP-----PERGAPS 1185
QY 1242 TFGKPTAENPEYLGIDVP 1260
DB 1186 IFKQ-PTAENAEYLRVAP 1203

```

RESULT 6  
 ERB4 HUMAN STANDARD; PRT: 1308 AA.  
 ID ERB4 HUMAN  
 AC 015303;  
 DT 15-DEC-1998 (rel. 37, last sequence update)  
 DT 15-JUN-2002 (rel. 41, last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)  
 DE (p180erbB4) (tyrosine kinase-type cell surface receptor HER4).  
 GN ERB4 OR HER4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).  
 RC TISSUE=Breast carcinoma;  
 RX MEDLINE=93189574; PubMed=8383326;  
 RA Plozman G.D., O'Leary J.-M., Whitney G.S., Green J.M., Carlton G.W.,  
 RA Foy L., Neubauer M.G., Shoyab M.,  
 RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the

RT epidermal growth factor receptor family.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=97476287; PubMed=9334263;  
 RA Elenius K., Corfas G., Paul S., Choi C.J., Rio C., Plozman G.D.,  
 RA Klagsbrun M.;  
 RT "A novel juxtamembrane domain isoform of HER4/erbB4. Isoform-specific  
 RT tissue distribution and differential processing in response to  
 RT phorbol ester.";  
 RL J. Biol. Chem. 272:26761-26768(1997).  
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-  
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND  
 CC NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.  
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB  
 CC RECEPTORS (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;  
 CC ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER  
 CC FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED  
 CC BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND  
 CC NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN  
 CC CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,  
 CC KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,  
 CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,  
 CC LUNG, SALIVARY GLAND, AND PANCREAS.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC -----  
 CC DR EMBL: L07868; AAB59446.1; .  
 CC DR HSSP: P11362; 1FGK.  
 CC DR Genew: HGNC:3432; ERBB4.  
 CC DR MIM: 600543; .  
 CC  
 CC InterPro: IPR000494; EGFR\_L\_domain.  
 CC InterPro: IPR000719; Euk\_Pkinase.  
 CC InterPro: IPR002174; Furin-like.  
 CC InterPro: IPR001245; Tyr\_Pkinase.  
 CC InterPro: IPR004019; YLP\_motif.  
 CC Pfam: PF00069; Pkinase; 1.  
 CC Pfam: PF00757; Furin-like; 1.  
 CC Pfam: PF01030; Recep\_L\_domain; 2.  
 CC Pfam: PF02757; YLP\_2.  
 CC Prodom: PD000001; Euk\_Pkinase; 1.  
 CC SMART: SMO0261; FU; 4.  
 CC SMART: SMO0219; TyKc; 1.  
 CC PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 CC PROSITE: PS00109; PROTEIN KINASE TYR; 1.  
 CC PROSITE: PS50011; PROTEIN KINASE DOM; 1.  
 CC Transmembrane: Glycoprotein; Multigene family; Receptor; Signal;  
 CC Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 CC Alternative splicing.  
 CC KW SIGNAL 1 25  
 CC FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.  
 CC FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).  
 CC FT TRASMEN 652 675 POTENTIAL.  
 CC FT DOMAIN 676 1308 CYTOPLASMIC (POTENTIAL).  
 CC FT DOMAIN 186 334 CYS-RICH.  
 CC FT DOMAIN 496 633 CYS-RICH.  
 CC FT DOMAIN 718 985 PROTEIN KINASE.

FT NP\_BIND 724 732 ATP (BY SIMILARITY).  
 FT BINDING 751 751 ATP (BY SIMILARITY).  
 FT ACT\_SITE 843 843  
 FT DISULFID 189 197 BY SIMILARITY.  
 FT DISULFID 193 205 BY SIMILARITY.  
 FT DISULFID 213 221 BY SIMILARITY.  
 FT DISULFID 217 229 BY SIMILARITY.  
 FT DISULFID 230 238 BY SIMILARITY.  
 FT DISULFID 234 246 BY SIMILARITY.  
 FT DISULFID 249 258 BY SIMILARITY.  
 FT DISULFID 262 269 BY SIMILARITY.  
 FT DISULFID 293 304 BY SIMILARITY.  
 FT DISULFID 308 323 BY SIMILARITY.  
 FT DISULFID 326 330 BY SIMILARITY.  
 FT DISULFID 350 352 BY SIMILARITY.  
 FT DISULFID 507 520 BY SIMILARITY.  
 FT DISULFID 532 532 BY SIMILARITY.  
 FT DISULFID 536 552 BY SIMILARITY.  
 FT DISULFID 555 569 BY SIMILARITY.  
 FT DISULFID 559 577 BY SIMILARITY.  
 FT DISULFID 580 589 BY SIMILARITY.  
 FT DISULFID 593 614 BY SIMILARITY.  
 FT DISULFID 617 625 BY SIMILARITY.  
 FT DISULFID 621 633 BY SIMILARITY.  
 FT MOD\_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARSPLIC 626 648 NGPTSHDCIYFMTGSHSTPYQHA -> ISSIEDCIGLMD  
 FT (IN ISOFORM JM-B).  
 SQ SEQUENCE 1308 AA; 146807 MW; 564A80985D6761 CRC64;  
 Query Match 42.1%; Score 2864.5; DB 1; Length 1308;  
 Best Local Similarity 44.2%; Pred. No. 2,8e-147;  
 Matches 598; Conservative 185; Mismatches 390; Indels 179; Gaps 29;  
 QY 9 WGIILLALPRGA-----STQVCTGTMTKRLRPSPELTHDMLRHYOGGOVGVGNLELY 64  
 DB 8 WTVVSLVLAAGTVQPSDSQVCAGTNKLSLSDEQYRALAKRYENCEVVGNIETIS 67  
 QY 65 LPTNASLSPLODIOEVGYVLAHNGOVPLQLRIVRGTQLFEDNYALAVLDNGDPLN 124  
 DB 68 IEHRDLSFLRSVREYVGYVLAHNGOVPLQLRIVRGTQLFEDNYALAVLDNGDPLN 127  
 QY 125 NTTPVYASRGRLQLNSTEILKGVYLQGNPOLCYODITLMDITHKNNQLLTLI 184  
 DB 128 NF-----GIQELKLNLTIELNGVYVDNKLCTADITHMODIVRNPMPSLTLV 178  
 QY 185 DTRSPRCHPCSPMGKSRWCSESSDCLSTRVACAGG-AACKGPLPTDCHEQCAAG 243  
 DB 179 STNGSGCGRCHHSCGTG-RWCGPTENHCQLTRVCAEQDGRGCPYVSDCCRRACAG 237  
 QY 244 CTGPKHSCLACIHPNHSIGCEALCALVTYNTDTRESMNPBGRRTFGASCTYACPPY 303  
 DB 238 CSBPKOTDCFACKNFNDGACVTCQQTFFVYNPTTQLBHNFAKTYGAFVCKKCPHNF 297  
 QY 304 LSTDVGSCTLVCPHNOEYTAEDGTORCEKSPCARVCGYGLMOYIKANSKFIQTELE 363  
 DB 298 V-VDSSCVYACSSKMEV-ENNGIMKRCCTDICKAODGIOTGSLMAGVTDSSNIDK 355  
 QY 364 FAGCKIFGSLATLPSPFDDDPASNTAPLOPELOVPELTLEITGLVLYSAMPDSLPDS 423

```

Db      356 FINCTKINGNLIFLVTGHCDDPYNAIEADPEKXNRTVREITGFLNTQSMFPMNTDFS 415
Qy      424 VEQNLQYIRGRILLHNGAVSLTLOGLGSLGRLSGLGSLALIHNTHLCPVHTYPMW 483
Db      416 VFSNLVTIGRVLVSGSLTLKQOGLTSLQFOSLKEISGNYITDNLSCYHTHTMT 475
Qy      484 QLEFNPHQALHTANRPEDECVCBGALACHQICARGHCMGPGPTCVNCSQFLRGOECVEE 543
Db      476 TLFSTINQRIYIRDNRAENCTAGMVCNHLCSGDCGCGPDCUSCRFSRGRITIES 535
Qy      544 CRVQLGPREYVNAHRLCPHRECPQ-QNGSYTCFGEBAQCAVCAHYKPPCVACBS 602
Db      536 CNLYDGEFREBNESISVCDDPCQEKXEDGLTCHGGPNCYKSHFGKGPNCVEKCPD 595
Qy      603 GVPKDLSPYMPKPEDEGACQCPINCTHSCVDLDKGC-----PAEQRASPL 651
Db      596 GLQGANF--IFKXADPDRECHPCHPCTGCGNCPHTSDCIYYPWHTSHLTPQAR-TPL 652
Qy      652 TSIVSANV-GILLVVVLGVVFGIILIKRQOKIRKRYTMRLLQETELVEPLTPSGAMPNQA 710
Db      653 --IAAGVIGGLFLLIVGLTFAYVVRKRSIK-KKRALRRL-ETELVEPLTPSGTAPNQA 708
Qy      711 QMRILKEFNFTVSFWLRVPKVSASHLETYYKGIWIDGKNVKPVAKYLRNTPSKAN 770
Db      709 QRLIKE-----TELKRVKVLGSGAGCTYKGIWVEGETVKIPVAKILNETGPKAN 762
Qy      771 KEILDEAYVAVGSSPYVSRLLGLTSTVOLVTLMPYCLLDHYENRGRLSGODLN 830
Db      763 VEFMDEALIMASMDPHLVRLGCLSPITQVLTQMPHSCLEVEYHKONIGSQLLIN 822
Qy      831 WCMQIAKMGVLEVRVLRDLARNVLSPPNHYKTTDGLARLDDIDTEYHAGGKY 890
Db      823 WCVQIAKMGVLEVRVLRDLARNVLSPPNHYKTTDGLARLDDIDTEYHAGGKY 882
Qy      891 PIKMALESILRRKFTQSDVWSYGVTVWELMTFGAKYDIPAREIPDLLEKGERLPP 950
Db      883 PIKMALECIHYKFTQSDVWSYGVTVWELMTFGAKYDIPAREIPDLLEKGERLPP 942
Qy      951 PICTIDYVMIMVWCMMDLSECRPRFRELVSFSPMADPPRFVYIQNED-LGASPLDST 1009
Db      943 PICTIDYVMIMVWCMMDLSECRPRFRELVSFSPMADPPRFVYIQNED-LGASPLDST 1002
Qy      1010 FYSLLEDMDMDGLVDAEYLVLPQCGFCPPAPAGAGVWHHRSSSTRSGGDLTLGL 1069
Db      1003 FPCNLDEDEDEMDMAEYLVP-QAENIRPP-----ITSNARLDSNRS-----EIGH 1050
Qy      1070 EPSEEEAPRS-----PLAP-SEGAGSDVPCDGLGMGA 1100
Db      1051 SPPPAYTPMSGNCFVYRDGFPAAEQGVSEYRAPTSITIPAPVACGATAEIFDSCNGT 1110
Qy      1101 AKGLQSLPTDPSPLQRYSEDPVPLPS-----ETDGVAVPLTSCPOEYVYNQDVR 1153
Db      1111 LKRVAVHVEDSSTORYSADPTVFAPERSPRGELDEGIMTMRDKPKKEYLNPAVE--- 1167
Qy      1154 QPSPREGPLPAARPAGATLERAKTSPGKNGVYKQVAFGAVENPEYLTPOGAAPOR 1213
Db      1168 -----ENPFVSR-----KMGDLQ-----ALDNPENYHNAAG----- 1194
Qy      1214 HPPPA-----FSPAFLMYVYDQDPPPERGA- 1238
Db      1195 -PKRADEVYNEPLVYNTFANTLGKAEYLLKNNILSPERKAKKAFDNPDIWNHSLPRSTL 1253
Qy      1239 -PSTFKGPT-----AENPEYL 1255
Db      1254 QHEDYLOEYSTKYFYKONGRIRPIVAENPEYL 1265

```

```

DT      15-JUN-2002 (Rel. 41, Last annotation update)
DE      Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).
GN      ERBB4 OR TYRO-2.
OS      Rattus norvegicus (Rat).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX      NCBI_TaxId=10116;
RN      [1]
RP      SOURCE FROM N.A.
RC      TISSUE=Heart;
RX      MEDLINE=98221155; PubMed=9553078;
RA      Zhao Y.-Y., Sawyer D.R., Ballig R.R., Opel D.J., Han X.,
RA      Marchionni M.A., Kelly R.A.;
RT      "Neuregulins promote survival and growth of cardiac myocytes.
RT      Persistence of ErbB2 and ErbB4 expression in neonatal and adult
RT      ventricular myocytes."
RL      J. Biol. Chem. 275:10261-10269(1998).
RN      [2]
RP      SOURCE OF 848-901 FROM N.A.
RC      TISSUE=Sciatic nerve;
RX      MEDLINE=9122560; PubMed=2025425;
RA      Lai C., Lemke G.;
RT      "An extended family of protein-tyrosine kinase genes differentially
RT      expressed in the vertebrate nervous system."
RL      Neuron 6:691-704(1991).
RN      [3]
RP      SEQUENCE OF 1031-1198 FROM N.A.
RC      STRAIN=Sprague-Dawley; TISSUE=Spinal cord;
RX      MEDLINE=97184212; PubMed=9030624;
RA      Carroll S.L., Miller M.L., Frommelt P.W., Kim S.S., Corbett J.A.;
RT      "Expression of neuregulins and their putative receptors, ErbB2 and
RT      ErbB3, is induced during Wallerian degeneration."
RL      J. Neurosci. 17:1642-1659(1997).
CC      -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-
CC      2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND
CC      NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.
CC      NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).
CC      -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC      tyrosine phosphate.
CC      -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB
CC      RECEPTORS (POTENTIAL).
CC      -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC      -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING
CC      NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS
CC      OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE
CC      RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, AND
CC      HEART.
CC      -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC      RESIDUES (BY SIMILARITY).
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -1- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC      or send an email to license@sdb-sdb.ch).
CC      -----
CC      EMBL: AF041838; AAD08899.1; -
CC      EMBL: U52531; AAC53051.1; -
CC      DR      HSSP: P11362; 1FGK.
CC      DR      InterPro: IPR000494; EGFR_L domain.
CC      DR      InterPro: IPR000719; Euk_Pkinase.
CC      DR      InterPro: IPR002174; Furin-like.
CC      DR      InterPro: IPR001245; Ty_Pkinase.
CC      DR      InterPro: IPR004019; YLP_motif.
CC      DR      Pfam: PF00757; Furin-like; 1.
CC      DR      Pfam: PF00069; Pkinase; 1.
CC      DR      Pfam: PF01030; Recep_L_domain; 2.
CC      DR      Pfam: PF02757; YLP_2_domain; 2.
CC      DR      PRINTS: PR00109; TYRKINASE.
CC      DR      Prodom: PD000001; Euk_Pkinase; 1.

```

RESULT 7  
 ERBB4\_RAT STANDARD; PRT; 1308 AA.  
 AC Q62956; Q922N7;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)





DB 1195 --PEKAEDEVNEPLVNTFTNALGNAEYMKNSLSYPEKAKAFDNDPDMNSLPERST 1252  
 QY 1239 --PPSTFKGTP-----NENPEYL 1255  
 DB 1253 LQHPDYLOEVSTKYFYKONGRIRPIVANEPEYL 1285

RESULT 8  
 XMRK\_XIPMA STANDARD; PRT; 1167 AA.  
 ID XMRK\_XIPMA  
 AC P13388;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).  
 GN XMRK OR 10.  
 OS Xiphophorus maculatus (Southern platyfish).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
 OX NCBI\_TaxID=8083;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90015140; PubMed=2797166;  
 RA Wiltbrodt J., Adam D., Malitschek B., Maueier W., Raulf F.,  
 RA Telling A., Robertson S.M., Schartl M.,  
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-  
 RT inducing Tu locus in Xiphophorus";  
 RL Nature 341:415-421(1989).  
 RN [2]  
 RP REVISION TO 515.  
 RA Schartl M.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X16891; CAA34770.2; -  
 DR PIR: S06142; S06142.  
 DR HSBP: P11362; IRGK.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR002290; Ser Thr\_Pkinase.  
 DR InterPro: IPR001245; Tyr\_Pkinase.  
 DR Pfam: PF00069; Pkinase\_1.  
 DR Pfam: PF00757; Furin-like\_1.  
 DR Pfam: PF01030; Recep\_L\_domain\_2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_Pkinase\_1.  
 DR SMART: SM00261; Fv\_5.  
 DR SMART: SM00220; S\_TKc\_1.  
 DR SMART: SM00219; TyKc\_1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP\_1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR\_1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM\_1.  
 KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;  
 KM Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1167 MELANOMA RECEPTOR PROTEIN-TYROSINE

FT DOMAIN 26 642 KINASE.  
 FT TRANSMEM 643 665 EXTRACELLULAR (POTENTIAL).  
 FT DOMAIN 710 977 POTENTIAL.  
 FT NP BIND 716 724 CYTOPLASMIC (POTENTIAL).  
 FT BINDING 743 743 PROTEIN KINASE.  
 FT ACT SITE 835 835 ATP (BY SIMILARITY).  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 220 228 BY SIMILARITY.  
 FT DISULFID 224 236 BY SIMILARITY.  
 FT DISULFID 237 245 BY SIMILARITY.  
 FT DISULFID 241 253 BY SIMILARITY.  
 FT DISULFID 256 265 BY SIMILARITY.  
 FT DISULFID 269 296 BY SIMILARITY.  
 FT DISULFID 300 311 BY SIMILARITY.  
 FT DISULFID 315 330 BY SIMILARITY.  
 FT DISULFID 333 337 BY SIMILARITY.  
 FT DISULFID 504 513 BY SIMILARITY.  
 FT DISULFID 508 521 BY SIMILARITY.  
 FT DISULFID 524 533 BY SIMILARITY.  
 FT DISULFID 537 553 BY SIMILARITY.  
 FT DISULFID 556 569 BY SIMILARITY.  
 FT DISULFID 560 577 BY SIMILARITY.  
 FT DISULFID 593 615 BY SIMILARITY.  
 FT DISULFID 618 626 BY SIMILARITY.  
 FT DISULFID 622 634 BY SIMILARITY.  
 FT CARBOHYD 144 144 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 201 201 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 356 356 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 365 365 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 398 398 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 417 417 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 501 501 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC... ) (POTENTIAL).  
 FT CARBOHYD 621 621 N-LINKED (GLCNAC... ) (POTENTIAL).  
 SQ SEQUENCE 1167 AA; 129934 MW; 479380749DC1D55A CRC64;

Query Match 38.1%; Score 2698.5; DB 1; Length 1167;  
 Best Local Similarity 44.2%; Pred. No. 1.5e-132;  
 Matches 564; Conservative 167; Mismatches 396; Indels 149; Gaps 31;

QY 4 AALCRWGLIALLPQCAST---QVCTGDMKRLPASPEHLDMLRLYQCGQVQGN 59  
 DB 8 AALLQ--LILVLSISRCSTDPDRKVCQGTSNQMTM---LDNHVLMKKMYGCGNVLEN 62  
 QY 60 LETVTPNVAASPLDIDIOEVGYLIANOVROVQLRLIVRGQLFEDNYALAVLDN 119  
 DB 63 LETTYQENQDLSFLOSIOEVGYLIANNEVSTIPLVNLRLRGQNLTEGNTLLWSEN 122  
 QY 120 GDPNNTPTVTGASPGGLRELQRLSTELLKGVLIQRNPOLCYODTILMKDIFKRNQL 179  
 DB 123 YOK-NPSSP--DVGVGKQLQSLNLTILSGVVKVSHRPLLCNVEITIMWDIVXTSNP 179  
 QY 180 ALTLIDTNRSRACHPESPMCKSRCEWGESSEDCOSTIRVCAGGC-ARKKGLPLDCCHE 238  
 DB 180 TNNLIPHAFERQCCQKDHGCVNGSCWAPRGHCQKFKLLCEQCNRRRGKRPIDCCNE 239  
 QY 239 OCAAGCTGPKSHSDCLAHFNHSGICEHLCPALVLYNTDFESMNPREGRYFGASCVTA 298  
 DB 240 HCAGGCTGPRATDCLACRFNDGCKDCCPKKIYDIASHQVNVNPNKIFYGACAYVE 299  
 QY 299 CRYNLTSTVGSCTLVCPLANOEVAEDSTQCEKCKRCAKAVCYGLGQMYKANSKFLG 358  
 DB 300 CFSNVYVE-GACVNSCSAGMLEVD-ENGKRSCKKCDGCPYCGDIGISL-SYTIAN 356  
 QY 359 ITEL-EFAGCKKIFSLAFLPESFDGDPASNTAPQBPOLQVETLEITETGLYISAMPD 417  
 DB 357 SINISFSNCTINDIILNRNPFEGDPHYKIGTMDEPLMLTLTKETGLYLVMMWPE 416  
 QY 418 SLPDLSVFQNLQVIRGLIHNGAYS-LTLQIGISWLGRLSRLGSGALIHNTHTLCF 476

```

Db 417 NMISLVSQNLLEIRGRTTFSRGSFVVAVQVRLHQLGRLSKVSGNVILKXLTQLRY 476
Qy 477 VHTVPMQDLFRNPQALHTANRBEDECVEGLACHOLCAGHCHMGPGPTCCVNCOSQFLR 536
Db 477 ANTIMRRLTFSESDSIEYDART-----ENQCNNECESDGCWGGPMTTCVSCLIHVR 529
Qy 537 GQCEVBCRVUQGLPREVYVNAHCLPCHPCECOPONGSTTCGPEADQCACAHKXDPFC 596
Db 530 GGRCVASCNLLQGEFREAOVDRCVQCHQECLEVTQDSTCTCGPAPANCSSAKAFQDPQC 589
Qy 597 VARCPGVPDLSTVYPIKPFPEDEGACPCPINCTHSCVDLDDKGPACRASPSTISVS 656
Db 590 IPRCHGLSGDGLT-IKYNADKMGQCCQCHQNCCTQSGSGSLGCGKGD-IVSHSLAVG 647
Qy 657 AVVGLLVVLGVVFGILLIKRQOKIRKTYRRLLOETELVEPLTPSGAMNQAKMILK 716
Db 648 LVSGLLIVLIVALLIVLLRRRRRIK-RKRTRCLLOEKELVEPLTPSGQAPNQAFLRLK 706
Qy 717 --EENNFTVSFMLRVPKVSASHLETYVKGIMIPDGENKIVALKVIRENTSPANKEL 774
Db 707 ETEFKKRV-----LGSAGFVYVKGIMNPGENIRIVALKVIRENTSPANKEL 758
Qy 775 DEAYVAVGVSPYVSRLLIGICTSTVQLVTLMPYGCILDHVRENGRLSQDILLNCMQ 834
Db 759 DEAYVAVGVSPYVSRLLIGICTSTVQLVTLMPYGCILDHVRENGRLSQDILLNCMQ 818
Qy 835 IAKGMSYEDRVLRHDLAARNVLYKSNHVKITDFGLRLDDETEYHADGGKVPILK 894
Db 819 IAKGMSYEDRVLRHDLAARNVLYKSNHVKITDFGLRLDDETEYHADGGKVPILK 878
Qy 895 MALESILRRFTHSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 954
Db 879 MALESILRRFTHSDVMSYGVYVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 938
Qy 955 IDVYVIMKCMWIDSECRPRELVEFSRMAPRQPRVIVQNEDLGASPLDSTFTRSL 1014
Db 939 IEVYVIMKCMWIDSECRPRELVEFSRMAPRQPRVIVQNEDLGASPLDSTFTRSL 995
Qy 1015 LEDDMGDLVAEEYLVPOGFCPPDPAFGAGVHHRSSSTRSGGGLTLLEPSEE 1074
Db 996 LSSDD--DVDADEVLLPKRI-----NRQS----- 1020
Qy 1075 EAPRPLAPSEGAGSDVDGLGMAKGLSLPHDPSLQRISEDPV-PLPSETPGY 1133
Db 1021 --EPCIPPTGH-----PVRENSITLRNISPQNALKEDLGH 1056
Qy 1134 VAPLTCSPQPEYVNPQVPRQP-----PSPRE-----GPLP-AARPAQATLERAKTUS 1180
Db 1057 -----EYVNPQSETSRSLSDIYNPVYEDLTDGKGVLSLSQAEATNPSRPPEYLN 1106
Qy 1181 PGKGVGVKDFAPGAVENPEYLTPOGGAAPQHPDPAFSPAFDNLVYWDODPPERGAP 1240
Db 1107 TNQNSL--PLVSSSSMDPDY--QAG-----YQAAF-----LPQTAL 1141
Qy 1241 STFKGTPTANPEYLG 1256
Db 1142 GNGMFLPAENLEYL 1157

```

```

CX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90083234; PubMed=268785;
RA Kraus M.H., Issing W., Miki T., Popescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors."
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9031132; PubMed=2164210;
RA Ploman G.D., Whitney G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene."
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RC TISSUE=Placenta;
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT tyrosine kinase."
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTK.
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC "This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch)."
CC
DR EMBL; M29366; AAA35790.1; -
DR EMBL; M34309; AAA35791.1; -
DR EMBL; S61953; AAB26935.1; -
DR PIR; A36223; A36223.
DR HSSP; P11362; 1PKK.
DR Genem; HGNC:3431; ERBB3.
DR MIM; 190151; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TYK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE NEG.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; FALSE NEG.
DR PROSITE; PS00111; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KM

```



RESULT 10  
 ERB3\_RAT STANDARD: PRT: 1339 AA.  
 ID Q62799; Q62955;  
 AC Q62799; Q62955;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)  
 DE (c-erbB3).  
 GN ERB3.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 OX NCBI\_TaxID=10116;  
 [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=96096535; PubMed=8522190;  
 RA Hellyer N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;  
 RT "Cloning of the rat ErbB3 cDNA and characterization of the  
 recombinant protein.";  
 RL Gene 165:279-284(1995).  
 [2]  
 RP REVISIONS TO 85, 513 AND 565.  
 RA Hellyer N.J., Koland J.G.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE OF 922-1097 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
 RX MEDLINE=97184212; PubMed=930624;  
 RA Carroll S.L., Miller M.L., Frennet P.W., Kim S.S., Corbett J.A.;  
 RT "Expression of neuroligins and their putative receptors, ErbB2 and  
 ErbB3, is induced during Wallerian degeneration.";  
 RL J. Neurosci. 17:1642-1659(1997).  
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = AMP + protein  
 tyrosine phosphate.  
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
 SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
 AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
 PHOSPHATIDYLINOSITOL 3-KINASE.  
 CC -1- SIMILARITY: BELONGS TO THE ERB RECEPTOR FAMILY.  
 CC -1- This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
 or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 CC -----  
 DR EMBL: U29339; AAC28498.2; -  
 DR EMBL: U29339; AAC28498.2; -  
 DR HSSP: P11362; 1PGR.  
 DR InterPro: IPR000494; EGFR\_L domain.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_Pkinase.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L domain; 2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR PRODOM: PD000001; Euk\_Pkinase; 1.  
 DR SMART: SM00261; FU; 5.  
 DR SMART: SM00219; TYRK; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; FALSE\_NEG.  
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.  
 DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;

KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
 FT SIGNAL 1 19  
 FT CHAIN 20 1339  
 FT DOMAIN 20 643  
 FT TRANSMEM 644 662  
 FT CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 183 259  
 FT DOMAIN 707 964  
 FT NP\_BIND 713 721  
 FT BINDING 740 740  
 FT ACT\_SITE 832 832  
 FT DISULFID 186 194  
 FT DISULFID 190 202  
 FT DISULFID 210 218  
 FT DISULFID 214 226  
 FT DISULFID 227 235  
 FT DISULFID 231 243  
 FT DISULFID 246 255  
 FT DISULFID 259 266  
 FT DISULFID 290 301  
 FT DISULFID 305 320  
 FT DISULFID 323 327  
 FT DISULFID 500 509  
 FT DISULFID 504 517  
 FT DISULFID 520 529  
 FT DISULFID 533 549  
 FT DISULFID 556 573  
 FT DISULFID 576 585  
 FT DISULFID 589 610  
 FT DISULFID 613 621  
 FT DISULFID 617 629  
 FT CARBOHYD 126 126  
 FT CARBOHYD 250 250  
 FT CARBOHYD 353 353  
 FT CARBOHYD 408 408  
 FT CARBOHYD 414 414  
 FT CARBOHYD 437 437  
 FT CARBOHYD 469 469  
 FT CARBOHYD 522 522  
 FT CARBOHYD 566 566  
 FT CARBOHYD 616 616  
 FT CONFLICT 1028 1028  
 FT SEQUENCE 1339 AA; 147545 MM; 0AA5F2402BBDF1E CRC64;  
 Query Match 33.1%; Score 2271.5; DB 1; Length 1339;  
 Best Local Similarity 40.0%; Pred. No. 2,1e-114;  
 Matches 517; Conservative 171; Mismatches 437; Indels 167; Gaps 37;  
 QY 3 LALCWMGULLALPRGAA---STVCTGTDMLRLPASPEHTLMDLRLHYGCGVVOGN 59  
 DB 7 LQVLC---FLSLARGSEKNSQAVCPGTLGLSTGADNOYQTLTYLKECEVVMGN 62  
 QY 60 LELTYLPNTASISFLDDIEVOGYVLIANOVQVLEQLRIYRGQTLEFDNYALVLN 119  
 DB 63 LEIVLGNHNDLSFLQWIREVTVGYLVANNEFSVLEPLMLRVYRGVYDGRKAFVW-- 120  
 QY 120 GDLNNTTPVTGSPGGLRELOLRSTLELKGVLIGRPQCYQOTILMKOIFHNQNL 179  
 DB 121 ---LNTNT---NSSALROLKFTQTLLELGGVYLEKDKCHMDITLMDIVRV-- 170  
 QY 180 ALTLIDTNRSRACHPCSPMCKGRGCESSSEDCOSLTRVCAAGC-ARCKGLPTDCHE 238  
 DB 171 GAEIVVKNNGANCPCHVECKG-RCMGPBPDDQILTKITICAPQCGRGCFGNPNCCHD 229  
 QY 239 QCAAGCTGPRHSCLCLAFNHSIGTELHCPALVTYNTTFESMPPEGRYTFGASCVTA 298  
 DB 230 ECAAGGSGGQDDTCFACRRFNDSGACVPCPELVNKLTFQLEPHTKTYOYGGCVAS 289  
 QY 299 CPYNYLSTVGSCTVLCPLHNOEVTAEDETQRCSEKSKCARCYGL--GMOYIKANSKF 356  
 DB 290 CPNRFV-VQGTFCVRCAPPDKWEVD-KHGLKNCEPCGGLCPKACSESTGSGSRVQTVDSSN 347  
 QY 357 IGITLEPFAKCKKIFGSLAFLEPSFGDPASNTAPLQEPQLQVFTLEBITGYLYISAMP 416

DB 348 ID-----GFVWCTKILNLDPLITGLVNDVPMHKIPALDPEKXANFRTREITGYLNIQSWP 403  
 QY 417 DSLPDLSPFONLQVIRGRILHNGAYS-ITLQIGISIMLIGRSIRETSGSLALIHNTHTIC 475  
 DB 404 PHEHNSVPSNLTITIGRSILYNGFSLLIKNLVSLGRSKELISAGRVYISANQQLC 463  
 QY 476 FVHTVPMDOLEFRPHQALLHTA-NRPEDECVSGELACHQICARGHMGPGPTQVCNCSOP 534  
 DB 464 YHSLNMTLRLLRGPSEERLDIKYDRPLAGECLAEGKCDPLCSSGGCGMGPGQCLSCRY 523  
 QY 535 LRQGEVBEGRVGLPBYVYNNRHLCPCHPEQPONGSVTCGFPADQCVACAKYKDP 594  
 DB 524 SRGACVTHCNFQGEPRFVHAQCPSCHECLPMEGSTCNGSSSDACARCAHFRDP 583  
 QY 595 FCVACPSGVKPDLSYMPIMKFPDEGACOPCPINCTHSC--VDLDDKCPAQRASPLT 652  
 DB 584 HCNVSCPHGILG--ANGPIKYPDAGNECRPHENCTQCGNSELQDCLGQAEVLMKSH 641  
 QY 653 SIYSANVGLLVVGLGVNGLIKRQOKR-KYTRRLLOETELVEPLTPSGAMPNQAQ 711  
 DB 642 LVIAVTVG--LAVILMLGGSFLYWRGRRIQNNRRARYLERGESIEPLDPS-EXANKYL 698  
 QY 712 MRLLKEFNFTVSFMRVPSASHL-ETYYKGIWIPDGENYKIPVAIVLRENTSPKAN 770  
 DB 699 ARIKETE-----LRKLKVLGSGVPGVYHKGIMPEGESIKIPVCIVIEDKSGROS 751  
 QY 771 KEILDAVYVAGVGSPPYSRLIGITCTSTVOLVQMPYGCILLDHVRENRGLSGQDLN 830  
 DB 752 QAVTDHMLAVGSLDHAIIVRLGLCGSSSLQVLTQYLPGLSLDHVKQRRETLPGQLLN 811  
 QY 831 WCVQIKKGSYLEDAVLRHDLAARNVLKSPHVTITPFGARLLDIDETVHADGCV 890  
 DB 812 WGVQIKKMYLEHSHVRDLALRNVLKSPQVADPVAADLLPDDCKLHSEAT 871  
 QY 891 PIKMALESILRRRFTHSDVWSYGVVWELMTFGAKPYDGIPIREIPDLKEGRLPQ 950  
 DB 872 PIKMALESIHFGKYTHQSDVWSYGVVWELMTFGAEPYAGRLAETPOLLEKGERLQ 931  
 QY 951 PICTIDVYIMKCMWIDECRRFELVSEFSRMAKDRQRVVIONEDLGASPLDSTF 1010  
 DB 932 QCTIDVYVWKCMWIDINIRPTFELANETFRMARDPRYLVIKRAS-GGTTP--PAA 988  
 QY 1011 YSLLEDDMDGLVDAAEYLVPOQCFCDPAPGAGMHHRRSSSTRSGGDTLTIG 1070  
 DB 989 BSVLTTEL-----QELLEPEL-----DLDLLE 1014  
 QY 1071 PSEE-----EAPRPLAPBSG-----AGSDVFDGDLGMAAK 1102  
 DB 1015 AEEGLATLSGALSUPTGLTRPGOSILSPSSGYMNMOSLGEACLDAAVAGREQ 1074  
 QY 1103 GLQSLPTHDSPFORSEDPVPLPEETDGV--APL-----TC-----SPQPE-- 1144  
 DB 1075 FERPLSLH-PIPRGR-----PASSSGHTTGSABELOEVKVSRSRSRSPRGD 1126  
 QY 1145 --YVQPDVRFQPPSPREGR-----LPAARPAATLERAKTLP--GRNGV-- 1187  
 DB 1127 SAYHSGRHSILTPVPLSPGLEEDNGVMPDTHLRGASSREGTILSSVGLGTE 1186  
 QY 1188 ---KQVFATGAVENREYLPQGAAPQHP 1216  
 DB 1187 EEDDED-----EYEVYNNRKRQSP-PRPP 1209

RESULT 11  
 EGFR\_DROME STANDARD; PRT: 1426 AA.  
 AC P04412: O61601: Q9W2G0: P81868:  
 DT 13-AUG-1998 (Rel. 05, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (Egfr)  
 (Gurken receptor) (Torpedo protein) (Drosophila relative of ERBB).

GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 OK NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).  
 RX MEDLINE=94350209; PubMed=8070664;  
 RA Clifford R., Schubach T.,  
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals  
 RT that several genetically defined classes of alleles cluster in  
 RT subdomains of the receptor protein."  
 RL Genetics 137:531-550 (1994).  
 RN [2]  
 RP REVISIONS.  
 RA Clifford R., Schubach T.,  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE=85124611; PubMed=2982499;  
 RA Llywen E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.,  
 RT "The Drosophila EGF receptor gene homolog: conservation of both  
 RT hormone binding and kinase domains."  
 RL Cell 40:559-607 (1985).  
 RN [4]  
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.  
 RC STRAIN=Oregon-R; TISSUE=Embryo;  
 RX MEDLINE=87002474; PubMed=3093080;  
 RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.,  
 RT "Alternative 5' exons and tissue-specific expression of the  
 RT Drosophila EGF receptor homolog transcripts."  
 RL Cell 46:1091-1101 (1986).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=99102120; PubMed=9882502;  
 RA Leeshkin A.M., Yu S.-Y., Katz J., Baker N.E.,  
 RT "Several levels of EGF receptor signaling during photoreceptor  
 RT specification in wild-type, Ellipse, and null mutant Drosophila."  
 RL Dev. Biol. 205:129-144 (1999).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I).  
 RC STRAIN=Berkley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hinkins R.A., Gale R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Branton R.C., Rogers Y.-H.C., Blazey R.C., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.A., Baer A., Baxendale J., Bayraktaroglu L., Baasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brodtier P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Doolan K., Dou P.L., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Foeller C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heilmann T.J., Hernandez J.R., Hoack J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jalaal M., Kalush F., Karpen G.H., Ke Z., Kemison V.A., Ketchum K.A.,  
 RA Kimel B.E., Kodira C.D., Kraft C., Kravitz S., Klug D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris T., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pauley J.M.,  
 RA Palazolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,

RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Slier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RL "The genome sequence of *Drosophila melanogaster*,"  
 RL Science 287:2185-2195(2000).  
 RP [7]  
 RP SEQUENCE OF 959-1078 FROM N.A.  
 RC STRAIN=Daekwanryeong;  
 RX MEDLINE=85137938; PubMed=2983232;  
 RA Wadsworth S.C., Vincent W.S. III, Biodeau-Wentworth D.,  
 RT "A *Drosophila* genomic sequence with homology to human epidermal  
 RT growth factor receptor.";  
 RL Nature 314:178-180(1998).  
 RN [8]  
 RP SEQUENCE OF 1133-1137, 1155-1167 AND 1209-1216 FROM N.A., AND MUTATION  
 RP ANALYSIS  
 RX MEDLINE=92038942; PubMed=1936959;  
 RA Raz E., Schejter E.D., Shilo B.Z.,  
 RT "Allelic complementation among DER/Tib alleles: implications for  
 RT the mechanism of signal transduction by receptor-tyrosine kinases.";  
 RL Genetics 129:191-201(1991).  
 RN [9]  
 RP REVIEW  
 RX MEDLINE=97248481; PubMed=9094709;  
 RA Perrimon N., Perkins L.A.,  
 RT "There must be 50 ways to rule the signal: the case of the *Drosophila*  
 RT EGF receptor.";  
 RL Cell 89:13-16(1997).  
 CC -1- FUNCTION: BINDS TO FOUR LIGANDS: SPITZ, GURKEN, VEIN AND ARGOS,  
 CC WHICH IS AN ANTAGONIST. TRANSDUCES THE SIGNAL THROUGH THE RAS-RAF-  
 CC MAPK PATHWAY. INVOLVED IN A MYRIAD OF DEVELOPMENTAL DECISIONS.  
 CC CRITICAL FOR THE PROLIFERATION OF IMAGINAL TISSUES, AND FOR THE  
 CC DETERMINATION OF BOTH THE ANTERO-POSTERIOR AND DORSO-VENTRAL  
 CC POLARITIES OF THE OOCYTE. IN THE EMBRYO, PLAYS A ROLE IN THE  
 CC ESTABLISHMENT OF VENTRAL CELL FATES, MAINTENANCE OF AMNIOGROSA  
 CC AND VENTRAL NEUROECTODERMAL CELLS, GERM BAND RETRACTION, CELL FATE  
 CC SPECIFICATION IN THE CENTRAL NERVOUS SYSTEM AND PRODUCTION OF  
 CC CUTICLE  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: ISOFORMS TYPE I AND TYPE II ARE TYPE I  
 CC MEMBRANE PROTEINS. ISOFORM TYPE III IS A TYPE III MEMBRANE  
 CC PROTEIN.  
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS: TYPE I (SHOWN HERE), TYPE II AND  
 CC TYPE III. ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- TISSUE SPECIFICITY: UBICITOUSLY EXPRESSED IN EMBRYOS. IN LARVAE,  
 CC UNIFORM EXPRESSION IS SEEN IN WING DISKS, GENITAL DISK, ANLAGEN OF  
 CC TESTIS AND OVARY, AND BRAIN CORTEX. IN EYE-ANTENNA DISK, HIGHEST  
 CC EXPRESSION IS ANTERIOR TO MORPHOGENETIC FURROW, LEVELS REMAIN HIGH  
 CC IN PHOTORECEPTOR PRECURSOR CELLS. THIS PATTERN IS REVERSED IN  
 CC POSTERIOR EYE DISK. IN ADULTS EXPRESSION IS HIGH IN BRAIN CORTEX  
 CC AND THORACIC AND ABDOMINAL GANGLIA.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; AF052754; AAC08536.1; JOINED.  
 DR EMBL; AF052754; AAC08535.1; JOINED.  
 DR EMBL; AF052752; AAC08535.1; JOINED.  
 DR EMBL; K03054; AAA51462.1; -.  
 DR EMBL; K03417; AAA51460.1; -.

DR EMBL; K03416; AAA50965.1; -.  
 DR EMBL; K03418; AAA51461.1; -.  
 DR EMBL; AF109077; AAD26134.1; -.  
 DR EMBL; AF109078; AAD26132.1; -.  
 DR EMBL; AF109082; AAD26132.1; JOINED.  
 DR EMBL; AF109078; AAD26133.1; -.  
 DR EMBL; AF109084; AAD26133.1; JOINED.  
 DR EMBL; AF109079; AAD26130.1; -.  
 DR EMBL; AF109081; AAD26130.1; JOINED.  
 DR EMBL; AF109079; AAD26131.1; -.  
 DR EMBL; AF109083; AAD26131.1; JOINED.  
 DR EMBL; AF109080; AAD26135.1; -.  
 DR EMBL; AE003454; AAF6732.1; -.  
 DR EMBL; X02293; CAA26157.1; -.  
 DR EMBL; X78920; CAA55523.1; -.  
 DR EMBL; X78918; CAA55521.1; -.  
 DR EMBL; X78919; CAA55522.1; -.  
 DR PIR; A00640; GQFE.  
 DR HSSP; P11362; IFCK.  
 DR FLYBASE; FBgn0003731; Egfr.  
 DR Interpro; IPR000494; EGFR\_L\_domain.  
 DR Interpro; IPR000719; Euk\_Pkinase.  
 DR Interpro; IPR002174; Furin-like.  
 DR Interpro; IPR001245; Tyr\_Pkinase.  
 DR Pfam; PF00069; pkinase.1.  
 DR Pfam; PF00757; Furin-like.1.  
 DR Pfam; PF01030; Recep\_L\_domain.2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR PRODOM; PD000001; Euk\_pkinase.1.  
 DR SMART; SM00261; FU\_7.  
 DR SMART; SM00219; Tyrc.1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP.1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR.1.  
 DR PROSITE; PS00011; PROTEIN KINASE DOM.1.  
 DR Transmembrane; Glycoprotein; Receptor; Phosphorylation; Transferase;  
 KW Tyrosine-protein kinase; ATP-binding; Signal; Alternative splicing;  
 KW Developmental protein.  
 KM  
 FT SIGNAL 1 30 POTENTIAL.  
 FT CHAIN 31 1426 EPIDERMAL GROWTH FACTOR RECEPTOR.  
 FT DOMAIN 31 868 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 869 889 POTENTIAL.  
 FT DOMAIN 890 1426 CYTOPLASMIC (POTENTIAL).  
 FT DOMAIN 938 1198 PROTEIN KINASE.  
 FT NP\_BIND 944 952 ATP (BY SIMILARITY).  
 FT BINDING 971 971 ATP (BY SIMILARITY).  
 FT ACT\_SITE 1063 1063 BY SIMILARITY.  
 FT MOD\_RES 902 902 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 Query Match 27.4%; Score 1879; DB 1; Length 1426;  
 Best Local Similarity 32.0%; Pred. No. 2.4e-93;  
 Matches 460; Conservative 183; Mismatches 436; Indels 358; Gaps 44;  
 QY 24 QVCTGTDMKRLPASPETHLDMRLHYGGCVVQGNLELYLPT-NASLSFLQDIOEVOG 82  
 DB 100 KICITKSRLLSPSSKHEHYRLDRYNTCTVDNLTMLTLPENLDSFLDNIREVTVG 159  
 QY 83 YTLIANVQVQVPLQRLIVGTQF-----EDNYALAVLNDGDLNNTTPTGSPGL 137  
 DB 160 YLLISHVDKRVKVPFKLQIRKTLFSLVSEKALV-----TSK 203  
 QY 138 RELQRLSTLEILKGVLLQRPOLCYQDTILMKDIFHKKNQALATLLIDNRSRACHPSP 197  
 DB 204 YLLELPDRLVDLVNGVGHNNVNLCHMTIQSEIVSNGTAYVYDFTAEERECPKHE 263  
 QY 198 MKSGRCNGESSEDDQSLRTVYAGGCA-RCKGLPLPDDCHEQCAACTGPKSHDCLAC 255  
 DB 264 SCTHG-CWGEGRKQKSKLTCSPOCGRGYGRKPCCHLFPAGGCTPTQKDCIAC 322  
 QY 256 LHFNSGICELCPALVTYNTDTPESMNPBGRYTFGASCYACPYNTLSDVGSCTLVG 315  
 DB 323 KNFPDEAVSKSCPCPMRYNPTTYLLEINPBGKAYGATVKECP-GHLLDNMGACVWSC 381  
 QY 316 PLHNGEVTAEDEGTQRCBKSKPCAEVYGLGMQYIKANSKFLGITEL-----EPAGCKK 369

```

Db 382 PDRKMDKGE-----CPCNGPCPKTC-----PGTVLHAAGNISFNCTV 422
Qy 370 IFGSLAFPSFDG--DPAQNTA-----PLQBPOLQVFTLEITGYLISMPDQLD 421
Db 423 IDGNIRILDTFGSGFDVYANYTMGPRIYFLDPERREVFSTVEITGYLIEGTHQOFNR 482
Qy 422 LSVQNQLVIRGRILHNGAV-SLTLOGSLISWGLRRLRLGSLALIHNTLCPVHTV 480
Db 483 LSTRNLLETIHGRQLMSFAALAIYKSSLYSLMRLLKQISSGVITQINRDLCTVSN 542
Qy 481 PWDLFNPHQALLHTANRPEDECVGGLACHQLCARGHWGPGPTQCVNCSQFLRGQEC 540
Db 543 RMPAIQKEPEOKVWVWENLRADLCERKNGTICSDQCNEDGCGAGTQCLTKKFNFGTC 602
Qy 541 VEEGRVLOGLPREVYVNRHCLPCHPECOPONGSVTCGPBAPDQVCAHAKDPFCVARC 600
Db 603 IADGVYSNAVK--FDRNRTKICHPBCR-----TCNGAGADHCQECVHVDGQHCVSEC 654
Qy 601 P-----SGVK-----PDL 608
Db 655 PKNKYNDRGVCRECHATCDGCTGPKDTIGIAGCTTNALINNDATVRCCLLKDCKPD- 713
Qy 609 SYPIWKF--PDEGACOP-----CPI-----NCTH-----632
Db 714 GY--FWEYVHPQEGSLKPLAGRAVCRKCHPLCELCTNYGHEQVCSKTHYKRBOCET 771
Qy 633 -----SC-----VDLDKG-----641
Db 772 ECPADHYTDEQRCEQRFHPECNCGCTPGADCKSCNFKLPANEGPVNSTMNCTS 831
Qy 642 -CPAEOR-----ASPLTS-----TSAVVGILLVVLGVVFGI 673
Db 832 KCPLEMHVNYOYTAIGPYCAASPPSSKITANDVMIRIIGAVLPTICLCV--T 889
Qy 674 LIRKROQKIRKYT--MRLLQETELVPLTPSGAMPQAOIRLKEKNFTVSFWLRVPK 731
Db 890 YICRQKQKAKKETVMTALISGCEDEPLRPSNIGANLCKRLIVKDE-----LRKG 942
Qy 732 V-SASHLETYKGIWIDGENVKIPVAIKVLRNTSPKANKELIDEXYVAVAGSPYVR 790
Db 943 VLMGAFGRVYKGVWVEGENVXIPVAKELLKSTGASESEEFREXYIVASEHVNLLK 1002
Qy 791 LIGICLTSTVQVLTQMPYGLLDHVENRGRIGSQDLMWCMQJAKMGVLEVDVLYR 850
Db 1003 LLAQVSSQMLLITQLPLGLDLYANNRDKIGSKALLMWSTQIAGKSYLEKRLVHR 1062
Qy 851 DLAARNVLSKSPNHVKITDFGLARLIDIDEVEHADGKVPIKMMALLESILRRFTHSD 910
Db 1063 DLAARNVLTQPSLVKITDFGLAKLSSDSENYAAGKMPIKMLALECIRNRFVFSKD 1122
Qy 911 VMSGVTVMLMTFGAPYDGIIPAREIPDLLEKGERLPDPICITIDVYMTVMCKMIDSE 970
Db 1123 VMAFGVTIMELTFGQPHENIPAKDIPDLIEVGLKLEOBEISLIDYCLTSCMHLDA 1182
Qy 971 CRPRFELVSEFSKMDRPFQVYVITQNEIDG--PASPDLSTFYRSLIED--DMGDLVD 1025
Db 1183 MRPFKQULTVFAEFARDPGRYALIPGDKTRLPB-----YTSQEKRLIKLAPTTD 1235
Qy 1026 AEEYLVPOQGFPCDPAFGAGKGVHRHRSSTRSGGGLITLLEPSEEFAP-----RS 1079
Db 1236 GSEAIAPDYLQPKARPS-----HRTDCT-----DEMPKLNRYCXD 1274
Qy 1080 PLAPSEGAGSDVDFDQ--DLGMAKAGLQSLPTHDPSPFLQYSEDPVLPSPSTDGVAP 1136
Db 1275 PSKNSSSTGDDERDSSAREVGVNLR-----LDLPVEDDYLANP 1313
Qy 1137 LTGSPQPEYVNOQDVRQPSPREBGLPAARPAQATLERAKTISPGKNGVAVKVFAGGA 1196
Db 1314 -TCQPGNNNNNNNN--NPNQNNMAAVGAAGV--DLIGVPVS 1351
Qy 1197 VENPEYL--TPQGAAPQPH-----PPAFSP-AEDNLYYWD 1230

```

```

Db 1352 VNPEYLNLAQTLGVGESPIPTQITIGIPWGGPGMVEYVPMWPGSEPTSSDHEYYND 1408
RESULT 12
ID ERBB ALV STANDARD; PRT; 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian leukosis virus.
OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
OX NCBI_Taxid=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85228222; PubMed=2988784;
RA Nielsen T.W., Maroney P.A., Goodwin R.G., Rottman P.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erbB activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated BGF receptor.",
RL Cell 41:719-726 (1985).
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC -1- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INJECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTER ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC -1- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: M10066; AAA48763.1; ALT_INTT.
DR PIR: A00643; TVCHLV.
DR PIR: B00643; TVFVLV.
DR HSSP: P13362; 1PGR.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00069; pkinase, 1.
DR PRINTS: PRO0109; TYRKINASE.
DR PRODOM: PD000001; Euk_pkinase, 1.
DR SMART: SM00219; TYRKc, 1.
DR PROSITE: PS00107; PROTEIN KINASE ATP, 1.
DR PROSITE: PS00109; PROTEIN KINASE TYR, 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM, 1.
DR Transferrase; tyrosine-protein kinase; ATP-binding; Oncogene;
KM glycoprotein; phosphorylation.
FT DOMAIN 132 399
FT NP BIND 138 146 ATP (BY SIMILARITY).
FT BINDING 165 165 ATP (BY SIMILARITY).
FT ACT SITE 257 257 BY SIMILARITY.
SQ SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;
Query Match 24.3%; Score 1668.5; DB 1; Length 634;
Best Local Similarity 50.2%; Pred. No. 1.8e-82;
Matches 358; Conservative 83; Mismatches 143; Indels 129; Gaps 18;
Qy 587 CAHYKDPFCVAVRCPGSGVYKPLSYPIWKFPEEGACOPCPINCTHSCVDDDDKCPAEQ 646
Db 3 CAHFDGHCYKACACAGVLDGNDTL-VKRYADANAVALCLCPNCTRGKSGELBCP--- 58
Qy 647 RASPLTSLVSAVV-GILLVVLGVVFGILIRKROQKIRKYTRRLLQETELVEPLTPSGA 705

```

```

Db      59 NGSKTPTSIAGVGGCLLVAVGIGLYLRRR-HIVKRTLRRLDRELEVPPLPSGE 117
Qy      706 MPNOQWRILKEFNFTVSFWLRVPKVSASHLETYKGIIPDGENVKIPVAIKVLRNT 765
Db      118 APNOAHRLIKE-----TEFKKVKVLGSGAFVYKGLIPEGEKVKIPVAIKELREAT 171
Qy      766 SPKANKELIDEAYMAGVSPYVSRLLGLCTSTVQVLTQMLPQYGLDHWENRRLGS 825
Db      172 SPKANKELIDEAYMASVDNPHVCRLLGLCTSTVQVLTQMLPQYGLDHWENRRLGS 231
Qy      826 QDLNMCQIAKGSYSLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVYA 885
Db      232 QYLLNMCVQIAKGMVLEERRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVYA 291
Qy      886 DGGKVPDKMALESILRRFTHOSDVWSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGE 945
Db      292 EGGKVPDKMALESILRRFTHOSDVWSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGE 351
Qy      946 RLPOPICTIDVYIMTKCMWIDSECRPFREELVSEFSRMAPDPORFVIO-NEDLGPAS 1004
Db      352 RLPOPICTIDVYIMTKCMWIDSECRPFREELVSEFSRMAPDPORFVIO-NEDLGPAS 411
Qy      1005 PLDSTFYRSLIEDDMGDLVDAEYLVPOQGFCCDPAPAGAGMHHRRSSSTRSGGD 1064
Db      412 PTDSEFYRSLIEDDMGDLVDAEYLVPOQGFCCDPAPAGAGMHHRRSSSTRSGGD 449
Qy      1065 LTGLPSESEAPRSPF-----APSEGAGSDPFDGDLGMAKAGLQSLPTHSDPLQRY 1119
Db      450 -----SKTPLLSSLSATSNNSATNCID-----RNGQHPHREDSFYORIS 489
Qy      1120 EDPTVPLPSET--DGYVAPLTCSPQEPYVNOQDPVQPPSPREGPLPAPAPAGATERAK 1177
Db      490 SDPTGNLEESIDGFL-----PAPEYVNO--LMPKPPS-----521
Qy      1178 TLSFGKNGVADV-----AFGQVNEPYLTPQGAAPRHPAPASPA 1222
Db      522 -TAVVQNOIYNNISLTAISKLPMSRYQNSHSTVDNPEYL-----NTNOSPILAKTV 572
Qy      1223 FDNLYWDO-----DPEE-----RGAPSTFGTPTAENPEYLGLDVP 1260
Db      573 FESSPYWIGSGNHQINLNDPQYQODFLPNETKPRGLKVALENPEYLRLVAAP 625

RESULT 13
ERBB_AVIER STANDARD; PRT; 604 AA.
AC P00535;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
GN V-ERBB.
OS Avian erythroblastosis virus (strain Es4).
OC Viruses; Retroviridae; Retroviridae; Avian type C retroviruses.
OX NCBI_TaxID=79685;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=H;
RA MEDLINE=84026539; PubMed=6313229;
RA Yamamoto T., Nishida T., Miyajima N., Kawai S., Ooi T., Toyoshima K.;
RT "The erbB gene of avian erythroblastosis virus is a member of the src
RL gene family.";
RL Cell 35:71-76 (1983).
RN [2]
RP SEQUENCE OF 1-152 FROM N.A.
RA MEDLINE=8422957; PubMed=6328658;
RA Desjard B., Henry C., Benaisa M., Biserte G., Claverie J.-M.,
RA Sauter S., Martin P., Stehelin D.;
RT "Sequencing the erbB gene of avian erythroblastosis virus reveals a
RL new type of oncogene.";
RL Science 224:1456-1459 (1984).
-i- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

```

```

CC      tyrosine phosphate.
CC      -1- DISEASE: THE V-ERBB ONCOGENE TRANSFORMS AVIAN FIBROBLASTS AND
CC      ERYTHROBLASTS IN CULTURE AND INDUCES SARCOMAS AND ERYTHROLEUKEMIAS
CC      IN CHICKENS.
CC      -1- MISCELLANEOUS: V-ERBB IS A TRUNCATED AND MUTATED VERSION OF THE
CC      RECEPTOR FOR EPIDERMAL GROWTH FACTOR.
CC      -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      EMBL; K02006; AAA42394.1; ALT_INIT.
CC      EMBL; K01216; AAA42400.1; -.
CC      DR PIR; A00644; TVYU.
CC      DR HSSP; P11362; 1FGK.
CC      DR InterPro; IPR000719; Euk_pkinase.
CC      DR InterPro; IPR001245; Tyr_pkinase.
CC      DR Pfam; PF00069; pkinase; 1.
CC      DR ProDom; PD000001; Euk_pkinase; 1.
CC      DR SMART; SM00219; Tyrc; 1.
CC      DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC      DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC      DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.
CC      KW transferase; tyrosine-protein kinase; ATP-binding; Oncogene;
CC      glycoprotein; phosphorylation.
CC      KW DOMAIN 132 399
CC      FT NP_BIND 138 146 ATP (BY SIMILARITY).
CC      FT BINDING 165 165 ATP (BY SIMILARITY).
CC      FT ACT_SITE 257 257 BY SIMILARITY.
CC      FT CONFLICT 29 29 R -> W (IN REF. 2).
CC      FT CONFLICT 140 140 S -> F (IN REF. 2).
CC      FT CONFLICT 146 146 I -> V (IN REF. 2).
CC      SQ SEQUENCE 604 AA; 67633 MW; 76BECDD06745D609 CRC64;

Query Match: 23.7%; Score 1622; DB 1; Length 604;
Best local similarity 50.0%; Pred. No. 5; 3e-80;
Matches 348; Conservative 80; Mismatches 136; Indels 133; Gaps 17;

Qy      587 CAHYKDPPTCVARCSGVKPDLSYPIKWPDEBQACQPCPINCSTHSCVLDKQCPAQ 646
Db      3 CAHFDIDGPHCVACAGVAGENDTL-VKRYADANAVCCQCPNCTRGCKPGLECF 58
Qy      647 BASPLISIVSAV-GILVVVGVGVGILIKERQOKIRKYTMRLLOETELVEPLTPSGA 705
Db      59 NGSKTPTSIAGVGGCLLVAVGIGLYLRRR-HIVKRTLRRLDRELEVPPLPSGE 117
Qy      706 MPNOQWRILKEFNFTVSFWLRVPKVSASHLETYKGIIPDGENVKIPVAIKVLRNT 765
Db      118 APNOAHRLIKE-----TEFKKVKVLGSGAFVYKGLIPEGEKVKIPVAIKELREAT 171
Qy      766 SPKANKELIDEAYMAGVSPYVSRLLGLCTSTVQVLTQMLPQYGLDHWENRRLGS 825
Db      172 SPKANKELIDEAYMASVDNPHVCRLLGLCTSTVQVLTQMLPQYGLDHWENRRLGS 231
Qy      826 QDLNMCQIAKGSYSLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVYA 885
Db      232 QYLLNMCVQIAKGMVLEERRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEVYA 291
Qy      886 DGGKVPDKMALESILRRFTHOSDVWSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGE 945
Db      292 EGGKVPDKMALESILRRFTHOSDVWSGYVTWELMTFGAKPYDGIIPAREIPDLLEKGE 351
Qy      946 RLPOPICTIDVYIMTKCMWIDSECRPFREELVSEFSRMAPDPORFVIO-NEDLGPAS 1004
Db      352 RLPOPICTIDVYIMTKCMWIDSECRPFREELVSEFSRMAPDPORFVIO-NEDLGPAS 411
Qy      1005 PLDSTFYRSLIEDDMGDLVDAEYLVPOQGFCCDPAPAGAGMHHRRSSSTRSGGD 1064

```



Db 412 PTDSKFEYRTLMEEEDMEDIYDADEVLPVHOGFP-----NSPST----- 449  
 Qy 1065 LTGLPEPSEBEARSP-----APSEGAGDVDPDGLGMAKGLSLPTHPSPLOQRY 1119  
 Db 450 -----SRTELSSLSLSTNNKATNCID-----RNGGHVREDSPVQKRS 489  
 Qy 1120 EDPVPLPSET--DGVAAPLTCPQPEYVYVQDPVDPQPSFREGPLPAAPAGATLERAK 1177  
 Db 490 SDPTGNFLEESIDGFL-----PAPEYVYVQ--LMPKKESTAM----- 524  
 Qy 1178 TLPSPGNVGVKDVAF-----GGAIVENPEYLTPOGAPQPHPPA 1218  
 Db 525 -----VQVQYVNTSLTALSILKLPMSRYQNSHSTAVDNPEYL-----NTQSP 568  
 Qy 1219 FSPAFDNLVYWDQPPERGAPSTFGKPTAENPEY 1254  
 Db 569 AKTVFESSPYWIGSGNHQ-----INTLNDPXY 594  
 RESULT 14  
 EGRF\_CHICK STANDARD; PRT; 703 AA.  
 ID EGRF\_CHICK  
 AC P13387;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)  
 DE (fragment).  
 GN EGRF.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88261272; PubMed=3260323;  
 RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,  
 RA Ulrich A., Vennstrom B., Schlessinger J., Givol D.;  
 RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,  
 RT expression in mouse cells, and differential binding of EGF and  
 RT transforming growth factor alpha.";  
 RL Mol. Cell. Biol. 8:1970-1978(1988).  
 CC -1- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,  
 CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND  
 CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC This Swiss-Prot entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: M20386; AAA48760.1; -  
 CC InterPro: IPR000494; EGRF\_L domain.  
 CC InterPro: IPR000719; Euk\_Pkinase.  
 CC InterPro: IPR002174; Furin-like.  
 CC InterPro: IPR001245; Tyr\_Pkinase.  
 CC Pfam: PF00757; Furin-like; 1.  
 CC Pfam: PF01030; Recep\_L domain; 2.  
 CC SMART: SM00261; Fu; 4.  
 CC PROSITE: PS00107; PROTEIN\_KINASE\_ATP; PARTIAL.  
 CC PROSITE: PS00109; PROTEIN\_KINASE\_TYR; PARTIAL.

DR PROSITE; PS0011; PROTEIN\_KINASE\_DOM; PARTIAL.  
 KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;  
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.  
 FT SIGNAL 1 30  
 FT CHAIN 31 >703  
 FT DOMAIN 31 654  
 FT TRANSMEM 655 667  
 FT DOMAIN 668 >703  
 FT DISULFID 197 206  
 FT DISULFID 201 214  
 FT DISULFID 222 230  
 FT DISULFID 226 238  
 FT DISULFID 239 247  
 FT DISULFID 243 255  
 FT DISULFID 258 267  
 FT DISULFID 271 298  
 FT DISULFID 302 314  
 FT DISULFID 318 333  
 FT DISULFID 336 340  
 FT DISULFID 513 522  
 FT DISULFID 517 530  
 FT DISULFID 533 542  
 FT DISULFID 546 562  
 FT DISULFID 565 581  
 FT DISULFID 569 589  
 FT DISULFID 592 601  
 FT DISULFID 605 627  
 FT DISULFID 630 638  
 FT DISULFID 634 646  
 FT CARBOHYD 134 134  
 FT CARBOHYD 190 190  
 FT CARBOHYD 200 200  
 FT CARBOHYD 359 359  
 FT CARBOHYD 368 368  
 FT CARBOHYD 420 420  
 FT CARBOHYD 573 573  
 FT CARBOHYD 578 578  
 FT CARBOHYD 613 613  
 FT CARBOHYD 633 633  
 FT CARBOHYD 648 648  
 FT NON\_TER 703  
 SQ SEQUENCE 703 AA; 77427 MW; AFE2DE11B735A690 CRC64;  
 Query Match 23.3%; Score 1595; DB 1; Length 703;  
 Best Local Similarity 44.6%; Pred. No. 1.8e-76;  
 Matches 316; Conservative 111; Mismatches 251; Indels 30; Gaps 14;  
 Qy 8 RWGLLALLPPGA-----STOVCTGDMRLRUPASPETHLDMRLHYOCGVQGNLE 61  
 Db 13 RGAATLVLLGLVALCSAVEEKVCOGTNNKLTQGHVEDHFTSLQRYNNCEVLSNLE 72  
 Qy 62 LTYLPTNASTFLDIOGVGVLIANQVRAVPLQRIYRGTOLEFDNYALAVLNDGD 121  
 Db 73 ITVEHNDLTFKLTQVAGVYVIALMVAVIPLENIQIIRGVLYXNSPALVLSNYH 132  
 Qy 122 PLNNTPTVPGASPGARELQSTLEIKSGVLIQRPOLCYOPTILKDFHKNQAL 181  
 Db 133 -MNTKQ-----GLRELPMKRLSEILNGSVKSNPKLGMNTVLMNDILDSRK-PL 182  
 Qy 182 TLID-TNRSRACHPCSPCKSRGWGESSDDCGSLRTVCAAGCA-RCKGPLPDDCHEQ 239  
 Db 183 TVLDFASNLSSCPKHPNCTEDHGWGAGENCQTLTKVICAQCCSGRGKVPSPDCCHQ 242  
 Qy 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYVNTDTPESMNPESRYFFGASCVTAC 299  
 Db 243 CAAGCTGRESDDCLACRFDPDARCTKTCPLVLYNPTTYQMDVNPBGKYSFGATVAREC 302  
 Qy 300 PYNVLSTDVSGCTLVCPHNGEYTAEDGTORCEKSCRCARVCYGLQYIKANSKFIGI 359  
 Db 303 PHNYVTDHSGCVASCSNTDYEY--EENGVRCKKCDGLSKVNCNGIGIGELKGLIS--INA 360  
 Qy 360 TELE-FACCKRIIFSLAFIPSPFDGDRASNTAPLQPELOVYFETLEETGLVLYISAMPDS 418

Db 361 TNIDSPKCTKINGVSLVLAFLGDAFTKTLPLDPKKLDVFRVKEISGFLLIQAMPDN 420

Qy 419 LPLDSVFQNTQVIRGRLHNGAVSLTQGLISMLGRSLRELGSGLALIHNTLCEVH 478

Db 421 ATDLVAFENLEIIRGTRKHQGVSLAVNKLKISLGRSLKEISDGLAIIMKKNLCYAD 480

Qy 479 TVPDLQFRRNHQALLHTARPEDECGEGLAGHQLCARGHCMGPGPTQCVCNVSQFLRG 538

Db 481 TMTKRSLEFATOSQKTKIIONRNKQDCTADRHVCDPLCSDVCGMFGFHCFSRFSRQK 540

Qy 539 ECVEECRVLQGLPREVYNAHCLPCHECOPONG--SVTCFGEADQVACAHYKDPF 595

Db 541 ECVAQCNILQGEPEPEFERSKCLPCHECCLVQNSTAVNTTCSGPGPHCMKAHFLDGFH 600

Qy 596 CVACPSGVKPEKDSYMPIMKPEDEGACOPPCINCHSCVDLDKCGCAPQASPLTSIV 655

Db 601 CVKACPGAVLGENDTL-VMKADANAVCQCHPCCTGCKGPGLEGSP--NSKTPSIA 656

Qy 656 SAVV-GILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTP 702

Db 657 AGVVGGLLCLVVGIGLYLRRR-HIVRKTRRLRLQERELVEPLTP 703

RESULT 15

ERBB\_AVIEU STANDARD; PRT; 540 AA.

ID ERBB\_AVIEU

AC P11273;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).

GN V-ERBB.

OS Avian erythroblastosis virus (strain ts167).

OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.

OX NCBI\_TaxID=103898;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=87064458; PubMed=2878364;

RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;

RT "A single amino acid substitution in v-erbB confers a thermolabile phenotype to ts167 avian erythroblastosis virus-transformed erythroid cells.";

RT Mol. Cell. Biol. 6:1751-1759(1986).

RL Mol. Cell. Biol. 6:1751-1759(1986).

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein

CC tyrosine phosphate.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL/ M13179; AAA42401.1; -

DR PIR/ A25231; TVFEVB.

DR HSSP/ P11362; IRGK.

DR InterPro/ IPR000719; Euk\_Pkinase.

DR InterPro/ IPR001245; Tyr\_Pkinase.

DR Pfam/ PF00069; Pkinase; 1.

DR ProDom/ PD000001; Euk\_Pkinase; 1.

DR SMART/ SM00219; Tyrc; 1.

DR PROSITE/ PS00107; PROTEIN KINASE ATP; 1.

DR PROSITE/ PS00109; PROTEIN\_KINASE\_TYR; 1.

DR PROSITE/ PS00111; PROTEIN\_KINASE\_DOM; 1.

DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;

KM Glycoprotein; Phosphorylation.

FT DOMAIN 132 399 PROTEIN KINASE.

FT NP\_BIND 138 146 ATP (BY SIMILARITY).

FT BINDING 165 165 ATP (BY SIMILARITY).

FT ACT\_SITE 257 257 BY SIMILARITY.

FT VARIANT 270 270 H -> D (IN THERMOLABILE V-ERBB).

SO SEQUENCE 540 AA; 60412 MW; 58532297AA068865D CRC64;

Query Match 22.6%; Score 1552; DB 1; Length 540;

Best Local Similarity 52.8%; Pred. No. 2.6e-76;

Matches 330; Conservative 72; Mismatches 127; Indels 96; Gaps 15;

Qy 587 CAHYDPPFCVRCSPGVKPDLSYMPIMKPEDEGACOPPCINCHSCVDLDKCGCAPQ 646

Db 3 CAHFIDGPHCVACAGVAVGENDTL-VMKYADANAVCQCHPCCTGCKGPGLEGSP 58

Qy 647 RASPLTSIVSAVV-GILLVVLGVVFGILIKRQOKIRKYTMRLLOETELVEPLTPSGA 705

Db 59 NSKTPSIAAGVVGGLLCLVVGIGLYLRRR-HIVRKTRRLRLQERELVEPLTPSGE 117

Qy 706 MENQGMRLKEFNNFTYSFMLRVPKVASHLETYKGMIPDGENVKIPVAIKYLRNT 765

Db 118 AFNQHLRLKE-----TEFKRVYLGVGAGFYKGMWPEGEKVTIPVAIKYLRNT 171

Qy 766 SPKANKELIDEAYVAVAGSPYVSRLLGLICTSTVQLVTQMLPYGCLLDHVENRGRLGS 825

Db 172 SPKANKELIDEAYVAVASVDNPHVCRLLGLICTSTVQLVTQMLPYGCLLDYIREKDNIGS 231

Qy 826 QDLWKCQIACGMSLEDEVRLVHDDLAARNLVKSPNVKITTDFGLARLIDDETEYHA 885

Db 232 QYLLWVCQIAGKMYLEERHMYHDLAARNLVTPQHVKITDFGLAKQGADEXEYHA 291

Qy 886 DGAQVPIKMALESILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLKEGE 945

Db 292 EGKVPPIKMALESILHRIYTHQSDVMSYGVTVWELMTFGSRPYDGIIPAREISVLEKE 351

Qy 946 RLPQPPICITDYYMIMVKCMWIDSECRPFRELVSEFSRMAADPPQFVITQ-NEDLGPAAS 1004

Db 352 RLPQPPICITDYYMIMVKCMWIDSECRPFRELVSEFSRMAADPPRYVITQCDDEMHLPSS 411

Qy 1005 PLDSFPEFRLIEDDMGDLVDAEYLVLPQGGFCDDPAPAGAGMWHRRRSSSTSGGSD 1064

Db 412 PTDSKFYRLMEEDMEDIVDAEYLVPHQGF-----NSRST----- 449

Qy 1065 LTLGLEPSEBEAPRSL-----APSEGASDVFDGDLGGAAGLQSLPTHDPSPLOYYS 1119

Db 450 -----SRFLSLSLATSNSATNCIDRNG-----H----- 476

Qy 1120 EDPTVPLPSETGVIAPLTCSPQPEVYNAQPDVRRPPSPREGPLLAARAGAT-LEKAT 1178

Db 477 -----PVREDGFL-----PAPEYVNO-LMPKKPSTAMVQNGIYNYISLTAISKLP 521

Qy 1179 LSPGKGVKDVAFAGAVENPEYL 1203

Db 522 DSRYN-----SHSTAVDNPEYL 539

Search completed: July 22, 2003, 09:18:43

Job time : 22.6928 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 47.0485 Seconds  
(without alignments)  
5522.503 Million cell updates/sec

Title: SEQ4-695-709-14

Perfect score: 6853  
Sequence: 1 MELALCRWGLLALPPGA.....TFKGTPTAENPEYLGLDVEV 1261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

1: SP\_ARCHAEA:\*  
2: SP\_BACTERIA:\*  
3: SP\_FUNGI:\*  
4: SP\_HUMAN:\*  
5: SP\_INVERTEBRATE:\*  
6: SP\_MAMMAL:\*  
7: SP\_MHC:\*  
8: SP\_ORGANELLE:\*  
9: SP\_PHAGE:\*  
10: SP\_PLANT:\*  
11: SP\_PROTOZOA:\*  
12: SP\_VIRUS:\*  
13: SP\_VERTEBRATE:\*  
14: SP\_UNCLASSIFIED:\*  
15: SP\_VIRUS:\*  
16: SP\_BACTERIAP:\*  
17: SP\_ARCHAEP:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6130	89.4	1259	6	018735 canis fam1
2	3063	44.7	1209	11	090X70 ratu mus nov
3	3063	44.3	1210	11	090X70 ratu mus nov
4	2650	38.7	1165	13	09YH40 xiphophorus
5	2612.5	38.1	1137	13	09W6F6 gallus gall
6	2226	32.5	1328	13	P79754 fugu rubrip
7	1959.5	28.6	1433	5	09B1H9 anopheles g
8	1871	27.3	419	4	09UK79 homo sapien
9	1739	25.4	357	11	08R2X1 mus musculu
10	1697.5	24.8	412	4	08WYV0 homo sapien
11	1635	23.9	729	15	086712 avian rous-
12	1637	23.9	567	15	086714 avian rous-
13	1575.5	23.0	962	15	064895 avian eryth
14	1567	22.9	545	15	085468 avian eryth
15	1506.5	22.0	655	11	09WVFS mus musculu
16	1490.5	21.7	643	11	09ERV6 mus musculu

17	1204	17.6	1193	5	09Y1X8 ephydactia f
18	1139.5	16.6	1368	5	023821 caenorhabdi
19	1129	16.5	1717	5	026566 schistosoma
20	1126	16.4	527	13	090836 gallus gall
21	1001.5	14.6	478	11	09B5E0 ratu mus nov
22	942.5	13.8	599	13	09PSH2 gallus gall
23	906	13.2	165	4	014256 homo sapien
24	887	12.9	176	11	0923V5 ratu mus nov
25	806.5	11.8	346	13	P11776 xiphophorus
26	778	11.4	435	5	08SZW1 drosophila
27	754.5	11.0	311	13	089162 xiphophorus
28	734	10.7	311	13	08BUD7 homo sapien
29	723	10.6	149	6	09B6G6 cycotolagus
30	705.5	10.3	1362	5	09PVZ4 xenopus lae
31	687	10.0	1671	5	09NUV5 biophalar
32	653.5	9.5	1368	13	08UM85 paratichthy
33	645	9.4	1418	13	093457 scophthalmu
34	640.5	9.3	1369	13	08UM86 paratichthy
35	628	9.2	1358	13	073798 xenopus lae
36	626.5	9.1	1472	5	09U5A8 bombux mori
37	611	8.9	1412	13	08UM84 paratichthy
38	601	8.8	1245	13	09YGH8 scopthalmu
39	597.5	8.7	1418	13	08UM83 paratichthy
40	587.5	8.6	2144	5	09QV94 drosophila
41	577	8.4	1371	11	09QVW4 ratu mus nov
42	572.5	8.4	1091	4	09UMQ4 homo sapien
43	564	8.2	987	11	091YMO mus musculu
44	562.5	8.2	1072	4	09BTB0 homo sapien
45	561	8.2	935	4	09EL35 homo sapien

## ALIGNMENTS

### RESULT 1

018735

ID 018735

AC 018735

DT 01-JAN-1998 (TREMBLrel. 05, Created)

DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)

DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DE ErbB-2.

OS Canis familiaris (Dog).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.

OX NCBI\_TaxID=9615;

RN [1]

RP SEQUENCE FROM N.A.

RA Yokota H.;

RT "cDNA cloning of erbB-2 from canine mammary gland";

RL Submitted (0CT-1997) to the EMBL/Genbank/DBJ databases.

DR EMBL; AB008451; EBA23127.1; ..

DR HSSP; P11362; 1FCR.

DR InterPro; IPR002048; EF-hand.

DR InterPro; IPR000494; EGFR L domain.

DR InterPro; IPR000719; Euk pkinase.

DR InterPro; IPR002174; Furin-like.

DR InterPro; IPR001245; Tyr\_pkinase.

DR InterPro; IPR004019; YLP\_motif.

DR Pfam; PF00757; Furin-like; 1.

DR Pfam; PF00069; pkinase; 1.

DR Pfam; PF01030; Recep\_L\_domain; 2.

DR Pfam; PF02757; YLP\_2; 1.

DR ProDom; PD000001; Euk\_pkinase; 1.

DR SMART; SM00261; FU; 3.

DR SMART; SM00215; TyrKc; 1.

DR PROSITE; PS00018; EF\_HAND; UNKNOWN\_1.

DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.

DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.

DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.

KW ATP-binding; Transferase; Tyrosine-protein kinase

SEQUENCE 1259 AA; 137989 MW; E3764D9C4ACD46 CRC64;

Query Match 89.4%; Score 6130; DB 6; Length 1259;  
 Best Local Similarity 89.6%; Pred. No. 0;  
 Matches 1135; Conservative 43; Mismatches 75; Indels 14; Gaps 4;

QY 1 MELALCRWGLLLALPPGAASIVCTGTDMKLELPASPEHLDMLHLYQGCQVQGNL 60  
 DB 1 MELAMCRWGLLLALPPGAASIVCTGTDMKLELPASPEHLDMLHLYQGCQVQGNL 60  
 QY 61 ELTYLPTNASLSPFODIOEVQGYVLIAMNOVROVPLRLIRIVGTQLFEDNYALAYLDNG 120  
 DB 61 ELTYLPTNASLSPFODIOEVQGYVLIAMNOVROVPLRLIRIVGTQLFEDNYALAYLDNG 120  
 QY 121 DPLNTTPVTGASGGRELRSLTEILKGVLIQRNPOLCYQDTILMKDIFHKONQLA 180  
 DB 121 DPLNTTPVTGASGGRELRSLTEILKGVLIQRNPOLCYQDTILMKDIFHKONQLA 180  
 QY 181 LTLIDTNRSRACHCSMCKGSRGMSSESDCSLTFTVAGGARGKPLPDCCHEOC 240  
 DB 181 LTLIDTNRSRACHCSMCKGSRGMSSESDCSLTFTVAGGARGKPLPDCCHEOC 240  
 QY 241 AAGCTGKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNREGRYTFGASCVTACP 300  
 DB 241 AAGCTGKHSDCLACLFHNSGICELHCPALVYNTDTFESMPNREGRYTFGASCVTACP 300  
 QY 301 YNYLSTDVGSCCTVCPINQEVTAEDGTORCEKSKRCARVCYGLMCIYKANSKFTGIT 360  
 DB 301 YNYLSTDVGSCCTVCPINQEVTAEDGTORCEKSKRCARVCYGLMCIYKANSKFTGIT 360  
 QY 361 ELRFAGCKIFGSLAFLEPESFDGDPASNTAPLOEQOVETLEETIIGYLIASMPDSL 420  
 DB 361 ELRFAGCKIFGSLAFLEPESFDGDPASNTAPLOEQOVETLEETIIGYLIASMPDSL 420  
 QY 421 DLVIFQNLQVIRGILINGAYSLTGLGIGSLGRLBRLGSLALIHNTLCPVHTY 480  
 DB 421 DLVIFQNLQVIRGILINGAYSLTGLGIGSLGRLBRLGSLALIHNTLCPVHTY 480  
 QY 481 PMDLPFRNPQALHTANREDECEVGEGLACHOLCAHGMGSPOTCCVNSQGLRQOEC 540  
 DB 481 PMDLPFRNPQALHTANREDECEVGEGLACHOLCAHGMGSPOTCCVNSQGLRQOEC 540  
 QY 541 VEECRVLOGLPREVYNARHCLPCHPECOPOGNSYTCGPPADOCVACAHYKDPFCV 600  
 DB 541 VEECRVLOGLPREVYNARHCLPCHPECOPOGNSYTCGPPADOCVACAHYKDPFCV 600  
 QY 601 PSQVPLDSTYMPIMKPFDEGACOPCPINCTHSCVDLDDGCPAEOBASLTSIVAVG 660  
 DB 601 PSQVPLDSTYMPIMKPFDEGACOPCPINCTHSCVDLDDGCPAEOBASLTSIVAVG 660  
 QY 660 PSQVPLDSTYMPIMKPFDEGACOPCPINCTHSCVDLDDGCPAEOBASLTSIVAVG 660  
 DB 660 PSQVPLDSTYMPIMKPFDEGACOPCPINCTHSCVDLDDGCPAEOBASLTSIVAVG 660  
 QY 721 FTVSPWLKRVKVSAS-HLETVYKGIWIPDGENYKIPALIVALENTPKANKILDEAY 779  
 DB 721 FTVSPWLKRVKVSAS-HLETVYKGIWIPDGENYKIPALIVALENTPKANKILDEAY 779  
 QY 779 FTVSPWLKRVKVSAS-HLETVYKGIWIPDGENYKIPALIVALENTPKANKILDEAY 779  
 DB 779 FTVSPWLKRVKVSAS-HLETVYKGIWIPDGENYKIPALIVALENTPKANKILDEAY 779  
 QY 840 SYLEDVRLVHRDLAARVVLKSPNHVKITPFGARLIDIDETEVHADGGVPIKMALES 899  
 DB 840 SYLEDVRLVHRDLAARVVLKSPNHVKITPFGARLIDIDETEVHADGGVPIKMALES 899  
 QY 899 SYLEDVRLVHRDLAARVVLKSPNHVKITPFGARLIDIDETEVHADGGVPIKMALES 899  
 DB 899 SYLEDVRLVHRDLAARVVLKSPNHVKITPFGARLIDIDETEVHADGGVPIKMALES 899  
 QY 900 ILRRRPTHOSDVSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPICTIDVYM 959  
 DB 900 ILRRRPTHOSDVSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPICTIDVYM 959  
 QY 959 ILRRRPTHOSDVSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPICTIDVYM 959  
 DB 959 ILRRRPTHOSDVSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGERLPPICTIDVYM 959  
 QY 960 IMKCMWIDSECRPRFELVSESRMARDPQRVYVIONELGASPLDSTFTFSLLEDD 1019  
 DB 960 IMKCMWIDSECRPRFELVSESRMARDPQRVYVIONELGASPLDSTFTFSLLEDD 1019  
 QY 1019 IMKCMWIDSECRPRFELVSESRMARDPQRVYVIONELGASPLDSTFTFSLLEDD 1019  
 DB 1019 IMKCMWIDSECRPRFELVSESRMARDPQRVYVIONELGASPLDSTFTFSLLEDD 1019  
 QY 1020 MGLDVAEEYLVPOQGFCDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEEBAPS 1079  
 DB 1020 MGLDVAEEYLVPOQGFCDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEEBAPS 1079

DB 1013 MODLVDAEEYLVPOQGFCDPAPAGAGVHHRRSSSTRSGGDLTLGLEPSEEBAPS 1072  
 QY 1080 PLAPSEBAGSDVFDGDLGMAAKGIQSLPTHDPSLQRYSEDPYPLPSETDGYAPLTC 1139  
 DB 1073 PLAPSEBAGSDVFDGDLGMAAKGIQSLPTHDPSLQRYSEDPYPLPSETDGYAPLTC 1132  
 QY 1140 SPOPEYVNOVDVPPSPREGLPAPAPAGATLER-----AKTLPKNGVVKDVFAFG 1194  
 DB 1133 SPOPEYVNOVDVPPSPREGLPAPAPAGATLER-----AKTLPKNGVVKDVFAFG 1192  
 QY 1193 SAVENBEYLAPRGRAAPQPPHPPAPSPAFDNLVYWDOPPERGAPSPFTFGTPTAENPEY 1252  
 DB 1255 LGLDVAV 1261  
 QY 1253 LGLDVAV 1259

RESULT 2  
 ID 090X70 PRELIMINARY; PRT: 1209 AA.  
 AC 090X70;  
 DT 01-MAY-2000 (TRENBLrel. 13, Created)  
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TRENBLrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor.  
 GN EGFR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.  
 ON NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER.  
 RX MEDLINE=90258688; PubMed=2342466;  
 RA Fetch L.A.; Harris U.; Raymond V.W.; Blaabard A.J.; Lee D.C.;  
 RA Earp H.S.;  
 RT "A truncated, secreted form of the epidermal growth factor receptor is  
 RT encoded by an alternatively spliced transcript in normal rat tissue."  
 RL Mol. Cell. Biol. 10:2973-2982(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER.  
 RA Fetch L.A.;  
 RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER.  
 RA Guttridge K.; Dawson T.L.; Earp H.S.;  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: M37394; AAF14008.1; -.  
 DR HSSP: P11362; IFGR.  
 DR InterPro: IPR000494; EGFR\_L domain.  
 DR InterPro: IPR000719; Euk pkinase.  
 DR InterPro: IPR002174; Furlin-like.  
 DR InterPro: IPR001245; Tyr pkinase.  
 DR Pfam: PR00757; Furlin-like; 1.  
 DR Pfam: PR00069; pkinase; 1.  
 DR Pfam: PR01030; Recep\_L domain; 2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PD000001; Euk pkinase; 1.  
 DR SMART: SM00261; Fu; 3.  
 DR SMART: SM00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00109; PROTEIN KINASE\_DOM; 1.  
 DR PROSITE: PS00109; PROTEIN KINASE\_TYR; 1.  
 KW ATP-binding; Receptor; Tyrosine-protein kinase.  
 SQ SEQUENCE 1209 AA; 134891 MW; 96FE7F6CC1B773 CRC64;

Query Match 44.7%; Score 3063; DB 11; Length 1209;  
 Best Local Similarity 49.1%; Pred. No. 1.5e-22;

Matches: 631; Conservative: 169; Mismatches: 362; Indels: 122; Gaps: 29;

QY 3 LALACRGLALLPEGA--ASTQVCTGTDMKRLPASPETHLMDLHYOGCVVGNIE 61  
 Db 15 LALLCAAG-----GALEKKVCOGTSNRLLTQGLTFEDHFLSLORMNCEVULNIE 66  
 QY 62 LITLPTNASSLFLQDIOEVOGYVLIHNOVROVPLORLIRVGTQJFEDNYALAVDNGD 121  
 Db 67 IIVORNYDLSFKTLQEVAGVLIATNTERIPLBNLOIRNNALEYNTVALAVLSN-- 124  
 QY 122 PLNNTPVTGASDGLRELQRLSTELTKGVLIGRNPOCQODTLMDIPKNNQAL 181  
 Db 125 -----YGTNKTGLBELPMRNDQELLIGAVFNNPILCMETIQMDIV QDVFLSN 175  
 QY 182 TLIDITNRS-RACHPCSPMCISRCWSESEDCSLTRVYACAGCA-RCKGPLPTDCHEQ 239  
 Db 176 MSMDVORHLTGCPKCDPSCPNGSCWGRGEENCOKLTKICACQCSRCRGRSPDCCHNQ 235  
 QY 240 CAGCTGPKRSDLLACHFNHSGICELHCPALVTYNTDPESMPNREGRTGASCVTAC 299  
 Db 236 CAGCTGPRESDCLVCHRFDEATCKDTCPLMLYPTTYQMDVNEGKYSFGATCVKXC 295  
 QY 300 PNYLSTDVGSCTLVPLHNOEYTAEDGTQRCCKSKPQARVYGLGMOYIKANSKFIGI 359  
 Db 296 PRVYVYTDHSGCVRACGPDYEV-EEDGVSCKCKDGPCKRVNGIGIEPK-DILSTIA 353  
 QY 360 TELE-FAGCKKITGSLAFLESPDGPASTAPLOEQLOVETLEITGYLYISMPDS 418  
 Db 354 TIKHFKYCAISGDHLIPVAFKGGSFRTPLDRELEILTYVEITGFLIQAMPEN 413  
 QY 419 LPDLSEYONLOVIRGRILHNGAYSLTLOGISMLGRSLREKSGSLALIHNTHLCEFH 478  
 Db 414 WDLHAFFENELIRGRTKOHGQSLAVGNTLSLRSKETSDSDVILISGRNLCYAN 473  
 QY 479 TVPMDOLFNRPHOALHTANRPDECVSGSLACHQICARHCWGPPTQCVNCSQFLRQ 538  
 Db 474 TIWKKLFGTPNOKTIMNRAKDKATNHYCNPLCSSEGCSEGPPTQCVSQNYSRGR 533  
 QY 539 ECVEEGRVIGLREYVNAHCLPCHPEGQPMGSGTCGFPFADQCVAAHKKDPFCA 598  
 Db 534 ECVDKCNILEGEBREVENSECTQCHPECTPOTMNTCTTGRGDNCKCAHYVDGHCVK 593  
 QY 599 RCPGSKPDLSTYMPWKFPDEGACQPCPINCTHSCVDLDDKCPAQRASP-LTSIYSA 657  
 Db 594 TCPSGIMGENNTL-VKFEADANNVCHLCHANCTYGCAGPLKCC--QDEGEPKIPSIANG 650  
 QY 658 VVGILLVVLGVVFGI-LIKRQOKIRKYMRLLOETELVELTSSGAMPNQOAKRIK 716  
 Db 651 IVGGILPLIV-VALGIGLEFMRROQVRKTLRLLOERELVEPLTSGGAPOAHRLIK 709  
 QY 717 EFNNTVSEFLRVPKYSASHLETYKGIWIPDGENYKIPVAIKVIRENTSPANKELIDE 776  
 Db 710 E-----TEFKIKIVLGSAGFVYKGLWIPBEGEKIIPVAIKELREASIPANKELIDE 763  
 QY 777 AYMAVAGSPYVSRLLIGICTLSTVQVLTQAMPYGLLDHYRENRGRSLQDILLNMCQIA 836  
 Db 764 AYMAVAVDPHVCRLIGICTLSTVQVLTQAMPYGLLDHYRENRGRSLQDILLNMCQIA 823  
 QY 837 KGMSTYEDVRLVLRDLAARVLYKSNHYKITDPGLARLLDDETEYHADGKVPKMA 896  
 Db 824 KGMNVLIEDRLVLRDLAARVLYKSNHYKITDPGLARLLDDETEYHADGKVPKMA 893  
 QY 897 LESILRRFTHOSDVMSYGVYTWELMTFSKAPYDGIIPAREIDPLEKGERLPOPICTID 956  
 Db 884 LESILHRIYTHOSDVMSYGVYTWELMTFSKAPYDGIIPASEIISILEKGRRLPOPICTID 943  
 QY 957 VVMIWYKCMWIDSECPRELVSESRAPORPVVIO-NEDGAPAPLUSTFRSL 1015  
 Db 944 VVMIWYKCMWIDSECPRELVSESRAPORPVVIO-NEDGAPAPLUSTFRSL 1003  
 QY 1016 EDDMDGLVDAEYVLPQGGFFCPDPAFGAGGVHHRHSSSTRSGGDLTLGLPSEEE 1075  
 Db 1004 EEDMDGLVDAEYVLPQGGFF-----NSPST----- 1030

QY 1076 APRSLPASEGASGVDFDQDLGMAKAGLQSLPTHPSPLORYSEDPTVP-PSER--DGY 1133  
 Db 1031 -SRTPLLSLSANSV-----SSTVACINNGSCRVKEDAFILQRYSDPTSVLTEDNIDTF 1085  
 QY 1134 VAPLTGSPQPEYVNOVDVRFPPSPRESPEPLPAARPAGTLERAKTLSSGKGVYVDVAF 1193  
 Db 1086 L-----VPEPEIND-SVSKRPAGSVQVFNHNOPLHP-----ABGRDLHYON--PH 1128  
 QY 1194 GGAIVENPEYL-TPQGAAPQHPHPAPFADNLYVMDQ-----DP-----PER 1236  
 Db 1129 SNAVSNPEYLTAAQ-----PCLSSGFSLSALWQKSHQMSLDNPDVYQDFFKE 1179  
 QY 1237 GAPPSTFKGTPAENPEYLGDPV 1260  
 Db 1180 AKPNGIFKG-PTAEVAYELRVAVP 1202

RESULT 3  
 Q9EP98  
 ID Q9EP98 PRELIMINARY; PRT, 1210 AA.  
 AC Q9EP98;  
 DT 01-MAR-2001 (Tremblrel, 16, Created)  
 DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)  
 DE Epidermal growth factor receptor isoform 1.  
 GN EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 RA Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,  
 RA Mable N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative Egr transcripts encoding truncated receptor  
 RT isoforms";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,  
 RA Schenl C., Pearse R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,  
 RA Mable N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative Egr transcripts encoding truncated receptor  
 RT isoforms";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF275366; AAC28045.1; JOINED.  
 DR EMBL; AF275364; AAC28045.1; JOINED.  
 DR EMBL; AF275365; AAC28045.1; JOINED.  
 DR EMBL; AF275367; AAC24386.1; -  
 DR HSSP; P11362; 1FGX  
 DR MGD; MGI:95294; Egr.  
 DR InterPro; IPR000345; CytC\_heme\_bind.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; EGFR\_L\_domain.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Kinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR PRINTS; PR00109; TyrKINASE.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SMO0261; FU; 5.  
 DR SMART; SMO0220; S\_TKC; 1.  
 DR SMART; SMO0219; TYKX; 1.

DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR ATP-binding; Receptor; Transferrase.  
 KM SEQUENCE 1210 AA; 134840 MW; 62CD021C9DE32E18 CRC64;

Query Match 44.3%; Score 3034; DB 11; Length 1210;  
 Best Local Similarity 48.6%; Pred. No. 2.4e-221;  
 Matches 622; Conservative 171; Mismatches 368; Indels 118; Gaps 26;

11 LLLALLPPGAA--STGYCTGTDKMLRLPASPEHLMRLHYGCGVQGNLELYLPTN 68  
 14 LLLALCAAGALEEKVKVCGQSTNRLLTGLTFEDHFLSLQRMNNCEVLGNLEITYVQGN 73  
 69 ASLSFLDIOGVQGVLIANNQVAVLQRLIVRGQLFEDYVYALAVDNGPLNTPR 128  
 74 YDSFLKTIQVAVGVLIANNVAVLQRLIVRGQLFEDYVYALAVDNGPLNTPR 124  
 129 VTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTLLMKDI---PKNNQLALTLI 184  
 125 -YGTNRRTGLRELPRNLOEILIGAVRFSNNPILCNMDTIQWRDIYQNVFMSNMSDL--- 180  
 185 DTRNSRACHPCSPKCKSRGCGWSSSDCQSLRTVCAGGCA-RCKGPLPTDCCHECAAG 243  
 181 -QSHPSSCPCKDPSCPGSCWGGEGENCQKLTKILCAQCCSHRCRGRSPDCCHNCCAG 239  
 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMNPPEGRTFGASCYACFPYNY 303  
 240 CTGPRESDCLVCKFOEATCKDTCPPLMLYNTTYQMDVNPBGKXSFQATCYKKCPRY 299  
 304 LSTDVSGCLVCPHLNNOVTAEDTQRCCKSKPCAVCYGLMKQYIKANSKITGITTELB 363  
 300 VVVDHSGCVACGPDYEV--EEGIRKCKCKDPCRCVKGIGIGEFK-DTLSINATNIK 357  
 364 -FAGCKKIFGSLAFLEPSFDGDDPASNTAPLOPEOLQVFETLEETGLYLSAMPDGLPI 422  
 358 HFYKCTAISGDLHLPLAFKGDSTRTPRDPRELLEIKVKEITGFLILQAMPDWTDL 417  
 423 SVFQNLQVIRGLHNGAVSLTLOGLISWLGRLSRELGSGLAIHNTLFCFHTVM 482  
 418 HAFENLEIIRGRTKOHQFSLAVVGLNITSLGRLSEISDGVIIISGNRLCYANTIMW 477  
 483 DDLFRNPHQALHTANPDECEVGEGLACHQLCARGHCWPGRTQCVNCSQPIRGQCEV 542  
 478 KKLFGTPNQKTKIMNNNAEKDCAVNVHVCNPLCSSEGCWPEPRDCVSNVSRGSECV 537  
 543 ECRVLQGLPREYVNAHCLPCHBECQPNQSVTCFGEADQCVAAHYKDPFCVACRS 602  
 538 KNILGEPREREVENSECIOCHPECLPQAMNITCTGRGPNGICQCAHYIDGPHCVTKCA 597  
 603 GVRPDLSTMPKWRPDEGACQPCINCHSCVDLDKCGPACQKASPLTISYSAVGLI 662  
 598 GINGENNTL-VMKYADANNVCHCHANCYTGACGPGIQGEVWPSGPKIPSIATGIVGL 656  
 663 LVVVLGVVFGI-LIKRQOKIKRYTMRLLQETELVEPLTPSGAMPNQAOMRLKEFNFF 721  
 657 LFLIV-VALGIGLFMRRHIVRKTLERLLQERELVEPLTPSEARPNQALRLTKE---- 711  
 722 TVSFVLAHVPKVASHLETYKGLTIPGGENVKIPVAIKVLRNTSPKANKLIDEAYVNA 781  
 712 --TEFKKIVAGSAGFTYKGLMIPGEEKYKIPVALKEKRETSPPANKEIIDEAYVNA 769  
 782 GVSGPYSRLLGLICTSTVQLVTLQMLPYGCLLDHVENRGLRSLQSLNNQCMQIAGMSY 841  
 770 SVDPNPHVCRLIGLICTSTVQLVTLQMLPYGCLLDHVENRGLRSLQSLNNQCMQIAGMSY 829  
 842 LEDVRLVHRDLAARNVLYKSPNHYKITDFGLARLLIDETRYHADGGKVPKIMMALESII 901  
 830 LEDRRLVHRDLAARNVLYKTPHVKITDFGLAKLGEKEKYHAEGKVPKIMMALESII 889  
 902 RRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIIDLEKEKERLPQPPICITIDVYMI 961

DB HRIYTHQSDVMSYGVTVWELMTFGSKPYDGIIPASDISISLEKGERLPQPPICITIDVYMI 949  
 QY 962 VKCMWIDSECPREBELVSEFSKMAKDPQRFVVIQ-NEDLGAPAPLSTFRSLLEDMD 1020  
 DB 950 VKCMWIDSECPREBELVSEFSKMAKDPQRFVVIQ-NEDLGAPAPLSTFRSLLEDMD 1009  
 QY 1021 GDLVDAEYLVPOQGFPCPDPAAGGVVHRHSSSTRSGGDLTGLSEEEAPRSP 1080  
 DB 1010 EDVVDADLYLTPQGF-----NSPST-----SRTF 1035  
 QY 1081 LAPSEAGSDVFDGLGCAAKGLQSLFTHDPSPLQRYSEDPTVPLPSET--DGVAPLT 1138  
 DB 1036 LLSLSLATS-N--NSTVACINRNGSCRVKEDAEFLQRYSSDPTGAVTEDNIDDAFL--- 1087  
 QY 1139 CSPQPEYVNOPDVRQPPSPREGPLPARPAGATLERAKTLSPGKGVVXDVFAFGAVE 1198  
 DB 1088 --FVDEYVNO-SVPRPAGSVQNPYHNPQHP-----APGRDLHYQN-PRSNVVG 1134  
 QY 1199 NPEYL-TPQGAAPQPHPPAFSPAFDNLVYWDQ-----DP-----PERGAPPS 1241  
 DB 1135 NPEYLNTAQ-----PTCLSSGFSNFPALMIQKSHQMSLDNPDYQDFFPKETKENG 1185  
 QY 1242 TPQCTPTAENPEYGLDVP 1260  
 DB 1186 IFKG-PTAENABYLRVAPP 1203

RESULT 4  
 Q9YH40 PRELIMINARY; PRT; 1165 AA.  
 AC Q9YH40;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Receptor tyrosine kinase proto-oncogene.  
 GN XMRK.  
 OS Xiphophorus xiphidium.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorphi; Acanthopterygii; Percormorpha; Atherinomorphi;  
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
 OX NCBI\_TaxID=8086;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RX MEDLINE=98241172; PubMed=9582016;  
 RA Dimitrijevic N., Winkler C., Wellbrock C., Gomez A., Duschl J.,  
 RA Altschmidt J., Scharf M.;  
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by  
 RT overexpression and mutational alterations.";  
 RL Oncogene 16:1681-1690(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RA Scharf M.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; US3471; AAD0500.2; -  
 DR HSSP; P1362; TRFK.  
 DR InterPro; IPR000345; Cyrc\_heme\_bind.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Kinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00261; Fu; 3.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
 DR PROSITE; PS00143; GRAM\_POS\_ANCHORING; UNKNOWN 1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS0011; PROTEIN KINASE DOM; 2.  
 DR PROSITE; PS00109; PROTEIN KINASE\_TYR; 1.  
 KW ATP-binding; Kinase; transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 1165 AA; 129614 MW; 7FEE38D771A74B CRC64;

Query Match 38.7%; Score 2650; DB 13; Length 1165;  
 Best Local Similarity 44.7%; Pred. No. 3,36-192;  
 Matches 574; Conservative 163; Mismatches 389; Indels 158; Gaps 33;

QY 1 MELALCRWGLLALLPRG-AAS-...QCTGDMRLRLPASETHLDMRLHYQCCQY 55  
 DB 4 LELLEL-...-LLLLLSIGRCSTDPKRCVQCGTSMQTM--LDNHYLLMKKMYSGCNV 56  
 QY 56 VQGNLELTYLPTNLSFLQDICEVGVYLIANQVAVPLRLRLRYRGTQLFEDNALA 115  
 DB 57 VLENLEITYQENDLSFLQICEVGVYLIANQVAVPLRLRLRYRGTQLFEDNALA 116  
 QY 116 VLDNGDPLNNTPYTGASPGGLRLQLSLTEILKGVLLIQRPQLCYQDTILMKDIFHK 175  
 DB 117 VMSNYOK-NPSSP--DYVQVGLKQLOLSNLTLEISGVKVSHPNELLNVEITIMWDIVDX 173  
 QY 176 NNQALTLIDNRSRACHPCSPCKSGSCWSESDCOSLTRVCAAGC-ARCKGPLEPTD 234  
 DB 174 TSNPTMLIPIAFERQCKDPCGCVNGSCWAPRGHCKTKLCAQCNRRCRGPPIID 233  
 QY 235 CCEHQAAGCTGPKHSDCLAFHNSGICELHALVYNTDTFESMPNREGRTFGAS 294  
 DB 234 CCEHQAAGCTGPKHSDCLAFHNSGICELHALVYNTDTFESMPNREGRTFGAS 293  
 QY 295 CVTACPNYLTSDVGSCTVCPHNOVYLAEDGQREKCSKPARCYGLG-...QYI 350  
 DB 294 CVKCFSPNYYVTE-GACVRSAGMLEVD-ENGRSKPCDGVCPKCDGIGSLSNTI 351  
 QY 351 KANSKFIQTELEFAGCKIFGSLAFLEPSEFGDPASNTAPLOEQVETLEITGYL 410  
 DB 352 AVNSTNIG-...-SSNCTKINGDIILNRNSPFGDPHYKIGMDEHMLNLTVEITGYL 407  
 QY 411 YISAWPDSLPDLSFQNLQVIRGLILNGAYS-LTQGLGSLWGLASLELGSGLALIH 469  
 DB 408 VIMWPMEMTSLSVFQNLLEIRGRTSRGFSFVAVVSHLQWGLSLEVSAGNVILK 467  
 QY 470 HNTLCEVAVPMPQLFNPQHQLHTANREDECEVSEGLACHOLCARGCWGPGPTQV 529  
 DB 468 NTPQLRYASTINMRLEFRSEDSQSLYEADRT-...-ENQCNNECEDDCWKGRTMVCV 520  
 QY 530 NCSQFLRGQSCVEECRYLQGLPREYVNAHRLCPHPCQFQNGSVTCFEPADQVACAH 589  
 DB 521 SCHVDRGRCVASCNMLQGEPRBAOVDRGVQCHQBELVQTDILTGYGEPANCSKCAH 580  
 QY 590 YKDPFCVACRPSGVKPLSLYMPIMKFPDEEGACOPRINCTHSCVDLDXGCPAEGRAS 649  
 DB 581 FQDPQCIIPRCPHMLGDBDTL-IMKADKMGQCPQCHQNTQSCSGPGLSCRGD-IVS 638  
 QY 650 PLTISAVVAILLVVLGVVGLIKRRQCKIRKYMRLLOETELVEPLTPSGAMPN 709  
 DB 639 HSSLAIVGLVSGILITVIALIIVLLRRRIK-RKRTIRRLQCEKLEVEPLTPSGQAPN 697  
 QY 710 AQMRILK--EFNNNTVSWFLVVPKVSASHLETYVKGIMWPGEVKIPVAILKYLRENTSP 767  
 DB 698 AFLRLKETEFKKORV-...-LGSAGFVTKGLMNPGENKIRIPVAILKYLRENTSP 749  
 QY 768 KANKEILDEAVYVAGVSPYVRLIGICLTSTVOLVTLQMPYGLLHVHVENRGLSQD 827  
 DB 750 KVNQEVLEAVYVAGVSPYVRLIGICLTSTVOLVTLQMPYGLLHVHVENRGLSQD 809  
 QY 828 LLNWCQIAKMSLEVDVRLVHRLAARNTVKSPPNVKITDFELALLDIDELHYADG 887  
 DB 810 LLNWCQIAKMSLEVDVRLVHRLAARNTVKSPPNVKITDFELALLDIDELHYADG 869  
 QY 888 GKVPYIKMALESIRRFTHQSDVSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKERT 947  
 DB 870 GKVPYIKMALESIRRFTHQSDVSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKERT 929

QY 948 POPPCTIDVYIMIMYKCMIMDECRPRFRELVSSEFSRMARDPQRFVYIQNEDLGPASPLD 1007  
 DB 930 POPPCTIDVYIMIMYKCMIMDECRPRFRELVSSEFSRMARDPQRFVYIQNEDLGPASPLD 986  
 QY 1008 STFYNSLEDDMDGLVDAAEYLVPOQGFCCDPAPAGAGVHHHRSSSTSGGDLTL 1067  
 DB 987 RLFSLSLSSDD--DVVDADDEVLT-...-RKRTIN-RQGS-... 1018  
 QY 1068 GLEPSEEARPSPLAPSEAGSDVDGDLGMAKGLQSLPTHDSPLQRYSEDPV-PL 1126  
 DB 1019 -...-EPCLIPNCH-...-PRENSIARLYSDPQNL 1047  
 QY 1127 PSETDGYVAPLTCSPOPEYVNOPDYRPOP-...-PSPRE-...-GLPL-ARPAAGATL 1173  
 DB 1048 EKDLDDH-...-EYVNGPGETSSRLSDIYVNPYEDLTDGMPVLSLSSQEAETNF 1097  
 QY 1174 ERAKTLSPKNGVAVDVFAGAVENPEVLTQCGAARPHPPAPAFEDLVYVDDQRP 1233  
 DB 1098 SREYVINTQNSL-...-PLVSSGSMDDPDY--QAG-...-YQAAF-...-L 1132  
 QY 1234 PERGAPSTFKGTPTENPEVLTGL 1257  
 DB 1133 PQTGALTGGMFLPAENLEVLTGL 1156

RESULT 5  
 Q9M6F6 PRELIMINARY; PRT; 1137 AA.  
 ID Q9M6F6;  
 AC Q9M6F6;  
 DT 01-NOV-1999 (TREMblrel. 12, Created)  
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)  
 DE Receptor tyrosine kinase (Fragment).  
 GN BRB84.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 ... (1) ...  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HINDRAIN;  
 RX MEDLINE=99263203; PubMed=10328884;  
 RA Dixon M., Lumsden A.,  
 RT "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in  
 RT embryonic chick hindbrain."  
 RL Mol. Cell. Neurosci. 13:237-258 (1999).  
 DR EMBL; AF121963; AAD31764.1; -.  
 DR HESP; P11362; 1FGK.  
 DR Interpro; IPR000494; EGFR\_L\_domain.  
 DR Interpro; IPR000719; Euk\_pkinase.  
 DR Interpro; IPR002174; Furin-like.  
 DR Interpro; IPR001368; TNFR\_c6.  
 DR Interpro; IPR001245; Tyr\_pkinase.  
 DR Interpro; IPR004019; YLP\_motif.  
 DR Pfam; PFG0757; Furin-like; 1.  
 DR Pfam; PFG0069; pkinase; 1.  
 DR Pfam; PFG01030; Recep\_L\_domain; 1.  
 DR Pfam; PFG02757; YLP\_2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; TyKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE\_TYR; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE\_TYR; 1.  
 KW Kinase; Tyrosine-protein kinase.  
 FT NON TER 1  
 SQ SEQUENCE 1137 AA; 127927 MW; 4D61436F87DC84F CRC64;

Query Match 38.1%; Score 2612.5; DB 13; Length 1137;

Best Local Similarity 45.6%; Pred. No. 2.2e-189; Indels 85; Gaps 27;  
Matches 523; Conservative 175; Mismatches 363;

```

QY 161 LCYODITLWMDIFKNNQALATLIDITNSRACHPCSPCKSGRCWGSSEDCOSLRTTVC 220
Db 3 LCFADTILHMDIVANPMASNFTLVPTNGSSGCGRCHSKCTG-FCMGFTENHCQTLTKTVC 61
QY 221 AGGC-ARCKGSLPDDCCHECCACAGTGPBKSDCLACHFNHSGICELHCPALTYNTDTF 279
Db 62 AECDDGRGYPYSDCCRCBCAGCGSGKDTDCACMKNFDSGCVTQCQTFYANFTTF 121
QY 280 ESMNPBGRYTFGASCVYACPYNLSTDVGSCTLVCPULHNOEYATDGTORCEKSCPKCA 339
Db 122 QLEHNNAKYTYGAFVCKKCPHFV-VDSSCVACAPSSKREV-EENGIMCKCKCTDICI 179
QY 340 RVCTGLGMQYIKANSKFIGITELE-FACCKKIFGSLAFLEPSFGDPASTAPLQEPOLQ 398
Db 180 KACDGIIGSL-VSAQTVDSSNIDKFNCTKINGNLFLVTGHDYHTIAALNPEKN 238
QY 399 VFETLEITGYLISAMPDLPDLSVFQNLQVIRGRILHNGVSLTQGLISWLGRL 458
Db 239 IFQYREITGYLNSQWPMNTDFRVSNTLVITGRLVSLILKQGITSLGQSL 298
QY 459 RELSGSLALHNNHLCFRTVPMDQFRNPHQALLHTANPEDECVGEGLAQOLCARG 518
Db 299 KQISAGNIYITDNSMLCYHTVNTSLFTSPQKTIHRNKAENCTADGVCNELCSD 358
QY 519 HGWGPTQCVNCSOFLRGCEVEECRYLQGLPREYVYARHCLCPHECOP-QNGSVTCF 577
Db 359 GCMGPTQCVNCSOFLRGCEVEECRYLQGLPREYVYARHCLCPHECOP-QNGSVTCF 577
QY 578 GPEADQVCAHAYKPPFCVAPCSGVKPPDLSYMPKWPDEBACOPPCINTCTHSCVDL 637
Db 419 GPGPDHCTKCFHFGDGRPCVCKCDGLOGANSF--IFKYADEDECHPCHECTQGRGP 476
QY 638 DDGC-----PACRASPLTSIVSAVV-GILLVVLGVVGLILRRROCKRKY 685
Db 477 ASHDIYPMTRQSTLPQHAR-TPL-IAAGVIGLFTIVVGLTFAVYARKSIK-KKR 532
QY 686 TMRRLQETELVEPLTPSGAMPNOAKMLKEFNFTVSFMLRVKPKVASHLETYVGIW 745
Db 533 ALRRL-ETELVEPLTPSGTAPNOQRIKE-----TELKRVKVLGSAFGTVYGIW 585
QY 746 IPDGENVKIPVALIKLRENTSPKANKELIDEAIVYAGGSYVSLGICLTSVQVLTQ 805
Db 566 VPEGETVAKIPVALIKLRENTSPKANKELIDEAIVYAGGSYVSLGICLTSVQVLTQ 805
QY 806 LMPYGLLDHYRENRGRGSDLLNMCMQIAKMSYLEDVRLVARDLAARVLYKSPNHV 865
Db 646 LMPYGLLDHYRENRGRGSDLLNMCMQIAKMSYLEDVRLVARDLAARVLYKSPNHV 865
QY 866 KITDFGLARLDIDETEHADGCKVPIKMMALLESILRRPFTHOSDWSYGVTVWELMTFG 925
Db 706 KITDFGLARLDIDETEHADGCKVPIKMMALLESILRRPFTHOSDWSYGVTVWELMTFG 925
QY 926 AKPYGIPAREIPDLLEKGERLPPICTIDVYIMVCMIMIDSECRPRELSEFSRM 985
Db 766 AKPYGIPAREIPDLLEKGERLPPICTIDVYIMVCMIMIDSECRPRELSEFSRM 985
QY 986 ARDPQRYVIONED-LGPASPLDSTFYSLLEDMDGDLVAEEVLYPQGFPCDDPAPG 1044
Db 866 ARDPQRYVIONED-LGPASPLDSTFYSLLEDMDGDLVAEEVLYPQGFPCDDPAPG 1044
QY 1045 AGGWVHHHRSSSTSGGGLTLGLEPSESEAPRS--PLAP-SEGAGSDVDDGLGMAA 1101
Db 885 SRTIDSRNNOVFVSDGYAAEQV-PMPYRAPGCIIPBAVYAGATATIEFTCCNGTL 943
QY 1102 KGLQSLPTHDSPPLQRYSEDPTVPLS-----ETDGYVALTCSPPQRYVNOBPVAPQ 1154
Db 944 KQVATLAKEDSDSTQRYGADPTVPIPERVIGELDEDGYMTPMRDKPTDVLNVEENPF 1003
QY 1155 PPSRREGPLPAA-RAGATLBERAKTLSPGKNCVVDVF-----AGCAVENEPEYITP 1205

```

```

Db 1004 VSRKNGDLQAVDNEPEYHN-----APNGQPKADEYNEPELYINTFANTLENAEL-- 1054
QY 1206 QGGAAPQPPPPAPSPAPDNLVYQDPPERPA--PSTFKCTP-----A 1249
Db 1055 -----KNLPEKAKAFDNPDMNHSLEPSTLQHPYLDSEYKTYKONGRIRPIVA 1108
QY 1250 ENPEYL 1255
Db 1109 ENPEYL 1114

```

## RESULT 6

```

ID P79754 PRELIMINARY; PRT, 1328 AA.
AC P79754;
DT 01-MAY-1997 (TEMBLrel. 03, Last sequence update)
DT 01-JUN-2002 (TEMBLrel. 21, Last annotation update)
DE Eub83.
OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
OC Acanthomorpha; Acanthopterygii; Perciformes; Tetraodontiformes;
OC Tetraodontidae; Takifugu.
OX NCBI_TaxID=91033;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9917347; PubMed=1007531;
RA Gellner K., Brenner S.;
RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu
RT rubripes."
RL Genome Res. 9:251-258(1999).
DR EMBL; AF056116; AAC34391.1; -.
DR HSSP; P11362; 1FGK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR Prodom; PD000001; Euk_kinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; Tyrc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW ATP-binding; Transferase.
SQ SEQUENCE 1328 AA; 148613 MW; A330309258B647B9 CRC64;

```

Query Match 32.5%; Score 2226; DB 13; Length 1328;

Best Local Similarity 39.4%; Pred. No. 6.5e-160; Indels 198; Gaps 33;

Matches 510; Conservative 158; Mismatches 427;

```

QY 9 WGLLLALPP--GAASTO-----VCTGTDKRLPLASPTHLDMLRLHYGCGVYQGNLEL 62
Db 4 WRLIMCVASRLAAASQGEAVCPGTQNGLSGSGQENQYLNKORYGCEIIMGNLEI 63
QY 63 TYPTNASLFLDIDIEVQGYULIANNQRYVQLRLRYGQLFEDNYALAVLDNGP 122
Db 64 TQIESWDFSEFLTLTEVGYVLIAMNHQELPLGGLRYIRGNLSLERRFALSFLIN-- 120
QY 123 LNNTPVTAASPGGLRELOLRSLTEILKGVLIQRPQVCYQDTILMKDIFHGNQALAT 182
Db 121 ----YRDS--PSGLNQLGIMNTLILDDGVQIINKYLRYGMYWYWRDI--RNNAPLE 173
QY 183 LIDTNSRACHPCSPCKSGRCWGSSEDCQSLRTVYACGC-ARCKGSLPDDCCHEQCA 241
Db 174 IQFNGBRGVCH--KSC-GNYCWPCKDCCQILTKTVCAQCCNDRCFGSPCCCHIEBA 229
QY 242 AGCTGPHSGCLACLHFNHSGICELHCPALVTYNTDTPFSMPYREGRYFGASCYVACY 301
Db 230 AGCKGSLDIDCFACRLFNDSGACVPCQPTLLYNKQTFQWETNPNAKYQGSICVQCFT 289

```



```

QY 302 NYLSTVSGCTLVCLPNOEVTADGTOR-CEKSKPCARVCYGLGMQYIKANSKFIGIT 360
DB 290 HFV-VDGSSCVSVCPKMEY--ERGSORCELCGSCPKVCEGTGA-----QROTDS 342
QY 361 ELE-FAGCKIFGSLALPESPDPPASNTAPLOPELOVPELEETITGLYISAMPDL 419
DB 343 NIDFINGTKIQSGLHPLVLTGILDDKRNVPFLDAKLVEFRVREITDILNOSMKEL 402
QY 420 PDLSVFONLQVIRIRILHNGAYSLLTQGLG:SMGLSLRELSELGSLALIHNTLCEVHT 479
DB 403 NDLVSFSLTTIGRSLFKRPSLMVMRIPTLISGLRSLRELISGSYIISQMLCHHT 462
QY 480 VPMQDLFRNPH-QALLTANRPEDECVGEGLAGHQLCARHCMPGPTQCVNCSQFLRG 538
DB 463 VNMQLFRGRVRANSUNSNRPMACVADGVCDPLSDSCWMPGPDQCLSRNYSRHG 522
QY 539 ECEVECRVLOGLPEEYVNAH-CLPCHPECOPQNGSVTCGFEADQCVACAHYKDPFCV 597
DB 523 TVVAGCHFNNSIPREFAGLNGVCVACHPECKPOTGKASCTGFGADECMACTKTRDGYCM 582
QY 598 ARCPGVKPDLSYMPIMKFPDEGACOPCPINCTHSCVDLDDKCPAEGASPLTISVA 657
DB 583 SSCPAGVN-DEKGLIEFKFPRBEGHCPCHQNTGCGSGPLNDC--LEAARLTISGQ 638
QY 658 VVGLLVVVGVP-----GLIKRQCKIRKRYTRRLLOETELVEPLTPSGAMNCA 710
DB 639 ITGIALGVPAGLICLVLPFLGMLYHNGLAIRRRGARRVLESGSEPELGP-GERKTV 697
QY 711 QMRILKEFNFTVFWLVPKVSASHLETYYKGIWIDGNNVKAIPVAKYLRNTPSKAN 770
DB 698 HARILKP-----SDLRKIFPLSGVGTGSKGFWIEGTEKIPVAKITIQSSSGRTF 751
QY 771 KEILDEAYVMAVGSPIVSRILGICLTSTVQVLTQMLPYGCLLDHRENGRIGSDLLN 830
DB 752 TEITDHLISGSDHPYIVRLIGCPGTCQLVLTQSSHSLHEHROHTSDDPORLLN 811
QY 831 WCMQIAKMSYLEDEVRILVHRDLAARNLVKSPNHVKITTDGLARLIDITEVHADGKV 890
DB 812 WCVQIAKMSYLEEHRVYHKNLAARNILKNDYQVQSDGVADLYPDDKITYSETKT 871
QY 891 PIKMALESILRRRTFHQSDVSVYGTVMELMTFGAKPYDGIAPAREIPDLLEKGERLP 950
DB 872 PIKMALESILFRRTYHQSDVSVYGTVMEMSGAPYASVQPEVPSYLEKGERLSQ 911
QY 951 PICTIDYVMIMVCKMWDSCORPRELVESEFSMARDDPGRFVLIQEDIGPASPLDST 1010
DB 932 AICTIDYVMIMVCKMWDENIRPTFKELASDFTRMARDPPRYLIRMEG----- 980
QY 1011 YRSLLDDMDGDLVDAEYLVPOGFCPPDAPAGGMVHRHRSSTRSGGDLTLGLE 1070
DB 981 -----EDSGMGEFL-----RGRSER--GLLEADLE 1003
QY 1071 PSEBEARSPLABEGAGSDVFDGDLGMC---AAKGLQSLPTHDPRLQ-----R 1117
DB 1004 EDEEE-----GLGDRFATPSLQPSPSWSTSPQSYNVMVTOUR 1042
QY 1118 YSEDPVPLPSETDGYVAPLTCSPQ- EYVNO-----PDVAPCPSPREG 1162
DB 1043 YD-----FAVSGGCHIGVLPMSBPYDITRQLMYQSRSLSVRTLPDRSAFRSSREAE 1096
QY 1163 L--PAAPAGATLERAKTILSPGNKVGVYVAFQGAVENTEYLTPOGGAAPORHPAPAS 1220
DB 1097 LCEBGAQACAGIFVR-----FGSERGN-----POG----- 1122
QY 1221 PAFDNLYWDDPPERGAPSTFGKGTAEENPE 1253
DB 1123 -----QORKUSTASSPSSFKTMADEDE 1146

```

RESULT 7  
Q9BIH9  
ID Q9BIH9  
AC Q9BIH9; PRELIMINARY; PRT; 1433 AA.

```

DT 01-JUN-2001 (Tremblrel, 17, Created)
DT 01-JUN-2001 (Tremblrel, 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel, 21, Last annotation update)
DE putative epidermal growth factor receptor (fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]_TaxID=7165;
RP SEQUENCE FROM N.A.
RC STRAIN=SLA;
RA Lycett G.J.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor";
RL Submitted (NOV-2000) to the EMBL/GenBank/DBD databases.
DR EMBL: AJ301655; CAC35008.1; -.
DR HSSP: P11362; 1FGK.
DR InterPro: IPR000345; CytC_heme_bind.
DR InterPro: IPR000494; EGFR_L_domain.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR InterPro: IPR001245; Tyr_pkinase.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR PRINTS: PR00103; Recep_L_domain; 2.
DR ProDom: PD000001; Euk_pkinase; 1.
DR SMART: SM00261; Fv; 7.
DR SMART: SM00220; S_TKc; 1.
DR SMART: SM00219; TyKc; 1.
DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE: PS00011; PROTEIN KINASE DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KM Receptor.
FT NON_TER
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRC64;

Query Match 28.6%; Score 1959.5; DB 5; Length 1433;
Best Local Similarity 31.9%; Pred No 1.3e-139;
Matches 464; Conservative 197; Mismatches 401; Indels 391; Gaps 38;

QY 26 CHTGDMKRLPASPETHLDMRLHYQCCQVQGNLELYLPTNALSFLQDIQEVQYVL 85
DB 1 CIGTNGMSVFPANREHYHNLDRYNTCTYVDGNLEITWQNTDNLNFIHQHREVTVYL 60
QY 86 IAHNGVQVPLORLIVGTOLF-----EDNYALAVLNDGDLNNTPTVTSFGGLREL 140
DB 61 ISLYDPQVILPRLQIRKRTTFKLNKMEAYGLFV-----SFSHMTL 104
QY 141 QURSLTEILKGVLLIQRPOLCYODTILMKDI-FHKNNOLATLIDTNSRACHPSPMC 199
DB 105 ELPALRLDIGSVGFNNYNLCHKMSIMWEILLAPQSMQVTFNPSSEERCPCHSC 164
QY 200 KSSRCWGESSEDCQSLTRTVACGGA--RCKAPLPTDCHCECAAGCTGPKNSDGLACH 257
DB 165 EVG-CWGEAHNQCQFSLKNCSPQSGRCFQPKRECHLFCAGGCTGPTQSDGLACKN 223
QY 256 FHNSGICEHLCPALTYNTDTPESMNPNEGRTFGASCVTACPRVYLSIDVSGCTLVCP 317
DB 224 FYDDGVCCQCECPMGIYNTPTNFWEPNPDGKAYATCYRKCP-EHLKNDGACVYRKCPK 282
QY 318 HNGEVTADGTORCEKSPCARVCYGLGMQYIKANSKFIGITELFAGCKKIFGSLAFL 377
DB 283 GMPQNSP-----CVPCKGVCPKTCPGEGI-----VHSDNIG---NYKDCITIEGSEIL 329
QY 378 PESPDDPASNT-----APLOEOLQVETLEELIINGLYISAMPDLSDLYSFONLQ 429
DB 330 DQSFQGFQGVYTNVSFGRYIKIDPRLEVFSVATEIGFTINQAHHPNFTTLNFRNLE 389

```

```

QY 430 VIRGRIILNGAY-SLTLOGLGISWLGRLSRLBSGLALIHNNHLCFVHTVPMDOLEFN 488
D 390 VVGRQLKENTLFAVYIVKISLKLSELKLRVNSGIVLENSDLFEVDDIMSEIKKS 449
QY 489 PHOALLTANPEDECEVGEGLACHOLCARHGWPGPTQCYNCSQPLRGOECVEECVLO 548
D 450 SDHEVWQKNENATECHEGEGECSCSKAGCWKGEGEQCLECNVYKXKCLDSCCK--- 506
QY 549 GLPREY-VNAHCLPCHRECOPONGSVTCFEPEDQVACAHYDPPFCYARCP----- 601
D 507 SLPRLYSVDSKTCGDCCHOECKD-----FCQPNEDNCGSCNNVXDRFCVAECPTTKHAM 561
QY 602 -----SGVKDLSYMPIWKEPD----- 618
D 562 NGTCINCHKTCVGRGPRDTIAPDGTISCDKAITGSDAKIERCLMKDESCPDGYSDYVL 621
QY 619 -EEG----- 621
D 622 QEEBPLKQLSGKAVCRKCHPRCKCTGYGPHQFCQECTGYKKGECDECDPODFYANEE 661
QY 622 -ACQPCPINCT-----HSCVDL-----DD-----KGCPEAQ----- 646
D 682 TRILPCGHCQCRGCHGLDDBHBECLNLFEGDGYDNATFTCVSNCPASHPYKRFQEA 741
QY 647 -----RASPLTISVAVGILLVVLGVVFGI---LIXRQCKIRKRYTM 687
D 742 GKIGPYCSADSMOGLRIEPOQYKIVGWSYMAILLICVFGIYFVFSHKKKDAVKM 801
QY 688 RRLLOETELVEPLTPSGAMPYQAOIRLIKFNFTVSMFVRPVASASHLETYVKGIMIP 747
D 802 TMLAGCEDSEPLRPSNVGRLTLRIKE-----AEIRGVYLGAGARGPRFKSVTMP 855
QY 748 DGENVKIPVAILKRENTSPYANKELIDEAVYVAGVSPYVSRLLGICLTSTVOLVTOIM 807
D 856 EGEVXKIPVAILKRENTSPYANKELIDEAVYVAGVSPYVSRLLGICLTSTVOLVTOIM 915
QY 808 PYGCLDHYHRENGRLSODLWCMQIAKMSYLEVRLVHRDLARNTLVYSPNNYKI 867
D 916 PLGCLDHYHRENGRLSODLWCMQIAKMSYLEVRLVHRDLARNTLVYSPNNYKI 975
QY 868 TDFGLARLLDIDETEHADGGKVDIKMMALLESILRRSFTHQSDVMSYGVTVWELMTFGAK 927
D 976 TVFGLAKLLDIDETEHADGGKVDIKMMALLESILRRSFTHQSDVMSYGVTVWELMTFGAK 1035
QY 928 PYDGIPIAREIPDLLEKGERLPOPICTIDYVMIMVKWMDSECRPFRELVSEFSMAR 987
D 1036 PYENVPAKDVPELIEIGHKLPQPDICSLDVYCIILSCWVLDADARPTFKOLATETFAEKAR 1095
QY 988 DPCRFVVIQNEDELGPASPLDSTFYRSLLEDDEDDMGDLV----- 1024
D 1096 DPCRFVVIQNEDELGPASPLDSTFYRSLLEDDEDDMGDLV----- 1146
QY 1025 -----DAEYLYVQOQFFCPDPAFGAGMVAHHRSSSTRSGGDLTLGLEPSEEFAPR 1078
D 1147 VPSTIAETDEYLCEKTRPSIMLPQPSA-----VBS6-DEMP6 1182
QY 1079 S-----PLAP-----SEGASDVFPGDGLGMAKAGLQSLPTHDEPLQRYSEDDPTVPLPSE 1129
D 1183 SLRYCKDPLKEDDETDHGEKVE-----GVGIR-----LNIPLD 1216
QY 1130 TDGIVAPLTCSPQPEYVNPQDVPQPSPREGLPAARPAAGATLERAKTSLPGKNGVYKD 1189
D 1217 EDDYLMPTCCSQ-----NQS-----TPG-----YMD 1237
QY 1190 VFAGGAVENPEYL-----TPGGAAPQHPHPPAFSPAFNMLVYMODPRERGAPR 1240
D 1238 LIGVPAVDNPEYLMGSTQALAGASMG--PHTPP-----PP 1274
QY 1241 STEFGTPTAENPE 1253
D 1275 NTPNGMPTHQHSQ 1287

```

```

RESULT 8
ID Q9UK79 PRELIMINARY; PRT: 419 AA.
AC Q9UK79,
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10465918;
RA Doherty J.K., Bond C., Jardim A., Adelstein J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
  autoinhibitor.";
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874 (1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelstein J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF177761; AAD56009.2; -.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-like.
DR Pfam; PF00757; Furin-like; 1.
DR SMART; SM00261; Fur; 1_domain; 1.
SQ SEQUENCE 419 AA; 45472 MW; FECLIB347E2D030C CRC64;

Query Match 27.3%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 1.1e-133;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 METALACRAGLLALLPRAASTQVCTGDMRLRPASETLDMRLHYQCCQVQVNL 60
D 1 METALACRAGLLALLPRAASTQVCTGDMRLRPASETLDMRLHYQCCQVQVNL 60
QY 61 ELTYLPTNLSLFLDIOGVQVYLIAHQVQVPLQRLIRVGTQLFEDNYALAVLNG 120
D 61 ELTYLPTNLSLFLDIOGVQVYLIAHQVQVPLQRLIRVGTQLFEDNYALAVLNG 120
QY 121 DPLANTPTVGTASPGGLRELQRLSLTEILKGVLIQRPQLCYDTILMKDIFHXNQJA 180
D 121 DPLANTPTVGTASPGGLRELQRLSLTEILKGVLIQRPQLCYDTILMKDIFHXNQJA 180
QY 181 LTLIDNRRACHPGSPMKGSRGSESESDCOSLTRTYCAGGACARCGPLPTDCCHQC 240
D 181 LTLIDNRRACHPGSPMKGSRGSESESDCOSLTRTYCAGGACARCGPLPTDCCHQC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMPNEGRYTFGASCVTACP 300
D 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTPESMPNEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPRLHNOETADGQRCGSKSPCARVYCYL 345
D 301 YNYLSTDVGSCTLVCPRLHNOETADGQRCGSKSPCARVYCYL 345
Db 301 YNYLSTDVGSCTLVCPRLHNOETADGQRCGSKSPCARVYCYL 345

RESULT 9
ID Q8R2X1 PRELIMINARY; PRT: 367 AA.
AC Q8R2X1,
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;

```

RN [1] SEQUENCE FROM N.A.  
 RP Strausberg R.;  
 RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC027080; AA027080.1;  
 KM Hypothetical protein.  
 SQ SEQUENCE 367 AA; 40163 MW; 0BE03395F9E101B0 CRC64;  
 Query Match 25.4%; Score 1739; DB 11; Length 367;  
 Best Local Similarity 88.0%; Pred. No. 9.5e-124;  
 Matches 323; Conservative 11; Mismatches 33; Indels 0; Gaps 0;  
 QY 895 MALESLIRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 954  
 DB 1 MALESLIRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 60  
 QY 955 IDVYIMVKCMIMDSCEPRFRELVESEFRMADDPORFVIONEDLGASPLDSTFYRSL 1014  
 DB 61 IDVYIMVKCMIMDSCEPRFRELVESEFRMADDPORFVIONEDLGASPLDSTFYRSL 120  
 QY 1015 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGGLTGLPSEE 1074  
 DB 121 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGGLTGLPSEE 180  
 QY 1075 EARRSLPABEGAGSDVFDGDLGMAAGLQSLPTHDPSPLOKRSSEDPVLPSETDGYV 1134  
 DB 181 EARRSLPABEGAGSDVFDGDLGMAAGLQSLPTHDPSPLOKRSSEDPVLPSETDGYV 240  
 QY 1135 APLTCSPOPEYVNPQDVPRPSPREGPLPAARPAATLERAATLSPGKNGVYKDYFAFG 1194  
 DB 241 APLTCSPOPEYVNPQDVPRPSPREGPLPAARPAATLERAATLSPGKNGVYKDYFAFG 300  
 QY 1195 GAVENPEYLTPOGGAAPQPPPPAFSPAFDNLTYWDQPPPRGAPPSFTFKTPTAENPEY 1254  
 DB 301 GAVENPEYLTPOGGAAPQPPPPAFSPAFDNLTYWDQPPPRGAPPSFTFKTPTAENPEY 360  
 QY 1255 LGLDVFPV 1261  
 DB 361 LGLDVFPV 367  
 RESULT 10  
 Q8WYV0 PRELIMINARY; PRT; 412 AA.  
 AC Q8WYV0;  
 DT 01-MAR-2002 (TReMBLrel. 20, Created)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)  
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)  
 DE Hypothetical 44.7 kDa protein.  
 GN PP3659.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 CX NCB1\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Huang Y., Zhou X.M., Zhang P.P., Jiang H.Q., Qin W.X., Zhao X.T.,  
 RA Wan D.F., Gu J.R.;  
 RT "Novel human cDNA clones with function of inhibiting cancer cell  
 RT growth."  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF318349; AAL55856.1;  
 DR InterPro: IPR002048; EF-hand.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00219; Tyrc; 1.  
 DR PROSITE: PS00018; EF\_HAND; UNKNOWN 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Hypothetical protein.

SQ SEQUENCE 412 AA; 44702 MW; 034397EFF3F2D2BC CRC64;  
 Query Match 24.8%; Score 1697.5; DB 4; Length 412;  
 Best Local Similarity 80.5%; Pred. No. 1.6e-120;  
 Matches 330; Conservative 5; Mismatches 22; Indels 53; Gaps 4;  
 QY 895 MALESLIRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 954  
 DB 1 MALESLIRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICT 60  
 QY 955 IDVYIMVKCMIMDSCEPRFRELVESEFRMADDPORFVIONEDLGASPLDSTFYRSL 1014  
 DB 61 IDVYIMVKCMIMDSCEPRFRELVESEFRMADDPORFVIONEDLGASPLDSTFYRSL 120  
 QY 1015 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGGLTGLPSEE 1074  
 DB 121 LEDDDMGDLVDAEYLVLPQGFPCPDPAAGAGVHHRRSSSTRSGGGLTGLPSEE 180  
 QY 1075 EARRSLPABEGAGSDVFDGDLGMAAGLQSLPTHDPSPLOKRSSEDPVLPSETDGYV 1134  
 DB 181 EARRSLPABEGAGSDVFDGDLGMAAGLQSLPTHDPSPLOKRSSEDPVLPSETDGYV 240  
 QY 1135 APLTCSPOPEYVNPQDVPRPSPREGPLPAARPAATLERAATLSPGKNGVYKDYFAFG 1194  
 DB 241 APLTCSPOPEYVNPQDVPRPSPREGPLPAARPAATLERAATLSPGKNGVYKDYFAFG 300  
 QY 1195 GAVENPEYLTPOGGAAPQPPPPAFSPAFDNLTYWDQPPPRGAPPSFTFKTPTAENPEY 1254  
 DB 301 GAVENPEYLTPOGGAAPQPPPPAFSPAFDNLTYWDQPPPRGAPPSFTFKTPTAENPEY 360  
 QY 1227 YWMD-QDPPR-----GAPSTFKGTPTAEN 1251  
 DB 361 YWMD-QDPPR-----GAPSTFKGTPTAEN 410  
 RESULT 11  
 Q86712 PRELIMINARY; PRT; 729 AA.  
 AC Q86712;  
 DT 01-NOV-1996 (TReMBLrel. 01, Created)  
 DT 01-NOV-1996 (TReMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)  
 DE Polypeptide.  
 GN POLYPEPTIDE.  
 OS Avian rous-associated virus type 1.  
 OC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.  
 CX NCB1\_TaxID=11950;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94203659; PubMed=8152791;  
 RA Vennstrom B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
 RA Johnson A., Beug H.;  
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
 RT different transforming capacities."  
 RL Oncogene 9:1307-1320(1994).  
 DR EMBL: S69372; AAC60725.1;  
 DR HSP: P03322; 1A6S.  
 DR InterPro: IPR007719; Euk\_pkinase.  
 DR InterPro: IPR004028; Retro\_M.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF02813; Retro\_M; 1.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00219; Tyrc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KW ATP-binding; Transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 729 AA; 80649 MW; 84DF6914E81D63 CRC64;  
 Query Match 23.9%; Score 1639; DB 15; Length 729;  
 Best Local Similarity 52.5%; Pred. No. 1e-115;

Matches 346; Conservative 79; Mismatches 126; Indels 108; Gaps 16;

QY 569 PQNSVTCFGEPAADCCVCAHAKDPFCVACGSPGKRDLSYMPIMKPEDEGACQCPPI 628  
 DB 141 PEETATPEKTGP--DHCKMCAHIDGPHCYKACPAVLGENDTL-VMKADANAVALCHP 197  
 QY 629 NCTHSCVDLDKGCACPAORASPLTSIVAVV-GILLVVLGVVFGILIKRQOKIRKTYM 687  
 DB 198 NCTRCCKGFGLEGCP--NGSKTPSIAGVVGGLLCLVVGIGLYLRR-HIVRKXTL 253

QY 668 RLLDETELVETPLTSSGAMPNOQMRILKEFNFTVSFWLVPKYSASHLETVYKGMIP 747  
 DB 254 RLLDETELVETPLTSSGAMPNOQMRILKEFNFTVSFWLVPKYSASHLETVYKGMIP 307

QY 748 DGENKIPVAIKVLRENTSPPKANKELDEAVYMAVGSPYVRLGICLTSTVQLTQM 807  
 DB 308 ESEKVKIPVAIKVLRENTSPPKANKELDEAVYMAVGSPYVRLGICLTSTVQLTQM 367

QY 808 PYGCLLDVRENRGRSGODLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKI 867  
 DB 368 PYGCLLDVRENRGRSGODLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKI 427

QY 868 TDFGLARLLDIDETRYHADGGKVPKIMMALESILRRFTHSDVMSYGVTVWELMTGAX 927  
 DB 428 TDFGLARLLDIDETRYHADGGKVPKIMMALESILRRFTHSDVMSYGVTVWELMTGAX 487

QY 928 PYDGIPIAEIPDLLEKGERLPQPICTIDVYIMVYKCMIDSECRPRELVSFSEMAR 987  
 DB 488 PYDGIPIAEIPDLLEKGERLPQPICTIDVYIMVYKCMIDSECRPRELVSFSEMAR 547

QY 988 DPQRFVIVQ-NEDLGASPLDSTFYRSLLEDMDMDVDAEYLVPOGFFCPRPAPGAG 1046  
 DB 548 DPQRFVIVQ-NEDLGASPLDSTFYRSLLEDMDMDVDAEYLVPOGFFCPRPAPGAG 598

QY 1047 GMVHHRHSSSTRSGGDLTLGLPESEEAAPRSP-APSEAGSDVDPDGLMGMAA 1101  
 DB 599 GMVHHRHSSSTRSGGDLTLGLPESEEAAPRSP-APSEAGSDVDPDGLMGMAA 625

QY 1102 KGLQSLPHDPSPLDQRYSEDPVPLPSET-DGYVAPLTCSQPEPYVNOQVPRPQPSPR 1159  
 DB 626 KGLQSLPHDPSPLDQRYSEDPVPLPSET-DGYVAPLTCSQPEPYVNOQVPRPQPSPR 675

QY 1160 EGPLPAAPAGATLERAKTLSPKNGVAVDV-APFGAVENPEYL 1203  
 DB 676 EGPLPAAPAGATLERAKTLSPKNGVAVDV-APFGAVENPEYL 715

RESULT 12

Q86714 PRELIMINARY; PRT; 567 AA.

AC Q86714;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)  
 DE V-erbB protein (Fragment).  
 GN V-erbB.  
 OS Avian rous-associated virus type 1.  
 OC Viruses; Retroviral viruses; Retroviridae; Alpharetrovirus.  
 NCBI\_TaxID=11950;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=94203659; PubMed=8152791;  
 RA Venström B., Raynoscheck C., Jansson L., Doederlein G., Lhotak V.,  
 RA Johnson A., Beug H.;  
 RT "Retroviral capture of c-erbB proto-oncogene sequences: rapid  
 RT evolution of distinct viral genomes carrying mutant v-erbB genes with  
 RT different transforming capacities.";  
 RL Oncogene 9:1307-1320(1994).  
 DR EMBL; S69372; AAC60727.1; .  
 DR HSSE; F11362; IFGK.  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00069; kinase; 1.

DR PRINTS; PR00109; TYRKINASE.  
 DR Prodom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00219; TyrcKc.1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00109; PROTEIN\_KINASE\_TYR; 1.  
 KM Tyrosine-protein kinase.  
 FT NON TER  
 FT 1  
 SQ SEQUENCE 567 AA; 63390 MW; C6D9CBA7ADF725E1 CRC64;

Query Match 23.9%; Score 1637; DB 15; Length 567;  
 Best Local Similarity 53.1%; Pred. No. 1e-115;  
 Matches 345; Conservative 77; Mismatches 120; Indels 108; Gaps 16;

QY 578 GPEADCCVCAHAKDPFCVACGSPGKRDLSYMPIMKPEDEGACQCPINCTHSCVDL 637  
 DB 1 GP--DHCKMCAHIDGPHCYKACPAVLGENDTL-VMKADANAVALCHPNTTRCKGP 57

QY 638 DDKGCPAORASPLTSIVAVV-GILLVVLGVVFGILIKRQOKIRKTYMRLDTEEL 696  
 DB 58 DDKGCPAORASPLTSIVAVV-GILLVVLGVVFGILIKRQOKIRKTYMRLDTEEL 113

QY 697 VEBLTSGAMPNOQMRILKEFNFTVSFWLVPKYSASHLETVYKGMIPDGENKIPV 756  
 DB 114 VEBLTSGAMPNOQMRILKEFNFTVSFWLVPKYSASHLETVYKGMIPDGENKIPV 167

QY 757 AIKVLRENTSPKANKELDEAVYMAVGSPYVRLGICLTSTVQLTQMIPYGCLLDHV 816  
 DB 168 AIKVLRENTSPKANKELDEAVYMAVGSPYVRLGICLTSTVQLTQMIPYGCLLDHV 227

QY 817 RERKRLGSDLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLL 876  
 DB 228 RERKRLGSDLLNMCQIAKMSYLEDVRLVHRDLAARNVLKSPNHVKITDFGLARLL 287

QY 877 DIBETRYHADGGKVPKIMMALESILRRFTHSDVMSYGVTVWELMTFGAKPYDGI 936  
 DB 288 DIBETRYHADGGKVPKIMMALESILRRFTHSDVMSYGVTVWELMTFGAKPYDGI 347

QY 937 IPDLLEKGERLPQPICTIDVYIMVYKCMIDSECRPRELVSFSEMARDPQRFVIVQ 996  
 DB 348 IPDLLEKGERLPQPICTIDVYIMVYKCMIDSECRPRELVSFSEMARDPQRFVIVQ 407

QY 997 NEDLGASPLDSTFYRSLLEDMDMDVDAEYLVPOGFFCPRPAPGAGVHHRHS 1055  
 DB 408 NEDLGASPLDSTFYRSLLEDMDMDVDAEYLVPOGFFCPRPAPGAGVHHRHS 451

QY 1056 SSTRSGGDLTLGLPESEEAAPRSP-APSEAGSDVDPDGLMGMAKGLQSLPTH 1110  
 DB 452 SSTRSGGDLTLGLPESEEAAPRSP-APSEAGSDVDPDGLMGMAKGLQSLPTH 485

QY 1111 DPSPFLQRYSEDPVPLPSET-DGYVAPLTCSQPEPYVNOQVPRPQPSPRPPLPAAP 1168  
 DB 486 DPSPFLQRYSEDPVPLPSET-DGYVAPLTCSQPEPYVNOQVPRPQPSPRPPLPAAP 526

QY 1169 AGATLERAKTLSPKNGVAVDV-APFGAVENPEYL 1203  
 DB 527 AGATLERAKTLSPKNGVAVDV-APFGAVENPEYL 566

RESULT 13

Q64895 PRELIMINARY; PRT; 962 AA.

AC Q64895;  
 DT 01-NOV-1996 (TREMBLrel. 01, Created)  
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Gag/v-erb-A/v-erb-B protein.  
 GN Gag/v-erb-A/v-erb-B.  
 OS Avian erythroblastosis virus.  
 OC Viruses; Retroviral viruses; Retroviridae; Avian type C retroviruses.  
 NCBI\_TaxID=11861;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA MEDLINE=90206603; PubMed=1969616;  
 RA Biskin A., Jackson J., Bishop J.M., McCarty D.J., Schatzman R.C.;  
 RT "Six amino acids from the retroviral gene gag greatly enhance the  
 RT transforming potential of the oncogene v-erb-B";  
 RL Oncogene 5:15-24 (1990).  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 DR EMBL: X52209; CA36459.1; JOINED.  
 DR EMBL: X52211; CA36459.1; JOINED.  
 DR HSSP: P10828; 2NLL.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR000535; Hormone\_rec\_119.  
 DR InterPro: IPR001723; Steroid\_receptor.  
 DR InterPro: IPR001245; Tyr\_kinase.  
 DR InterPro: IPR001628; Znf\_C4steroid.  
 DR Pfam: PF00104; hormone\_rec.1.  
 DR Pfam: PF00069; pkinase.1.  
 DR Pfam: PF00105; zfc4.1.  
 DR PRINTS: PR00398; STRDHOHOMNER.  
 DR PRINTS: PR00047; STROIDFINGER.  
 DR ProDom: PD000001; Euk\_pkinase.1.  
 DR ProDom: PD000035; Znf\_C4steroid.1.  
 DR SMART: SM00430; HOL1.1.  
 DR SMART: SM00219; TYK1.1.  
 DR SMART: SM00399; Znf\_C4.1.  
 DR PROSITE: PS00031; NUCLEAR RECEPTOR.1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP.1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM.1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR.1.  
 DR ATP-binding; DNA-binding; Nuclear protein; Receptor;  
 KW Transcription regulation; Tyrosine-protein kinase;  
 KW Zinc-finger.  
 SQ SEQUENCE 962 AA; 108320 MW; 3C5AED791E4B50CE CR664;

Query Match 23.0%; Score 1575.5; DB 15; Length 962;  
 Best Local Similarity 49.7%; Pred. No. 1,1e-110;  
 Matches 347; Conservative 76; Mismatches 150; Indels 125; Gaps 19;

541 VECCRYLQGLPRE-YVNAR-HCLP-----CHPEQ 568  
 354 IEKQESYLLAFEHYINRKHNIPIHWSKILMVAOLRMIGAVHNRFLMKYECETELS 413  
 569 PONGSVTCGPEADQVCAHYKDPFCVACRCPSPGLSYMPIKFPDEGACOPCP1 628  
 414 PGE-----VGP--DHCKCAHFIDGPHCVACPAVIGENDTL-VMKYADANAVCQLGHP 465  
 629 NCTHSCVDLDKCGPABQASPLTISVAVV-GILLVVLGVVFGILIKRQOKIRKXTM 687  
 466 NCTRGCKGPELEGP---NGSKTPSIAGVGGLLCLLVVVGIGLYLRR-HIVKRTL 521  
 688 RLLOETELVEPLTPSGAMNQAMRILKEFNFTVSFMLRVKVSASHLETYKGIWIP 747  
 522 RLLORELEVEPLTPSGEAPNQAHRLIKE-----TEFKYVVLGFGAFGLTYKGLWIP 575  
 748 DGNVNIPIVAIKLRENTSPKANKELIDEAYVAVGSPYVSLGLICLTSTVOLVQLM 807  
 576 EGKVTIPVAKIKLRATSPKANKELIDEAYVAVSDNPHVCGLLGICLTSTVOLVQLM 635  
 808 PYGCLLDHVENGRGSGDILNMCQIAKGSYLEDVRLVHRDLAARVNLVKSPPHVKI 867  
 636 PYGCLLDYIREHNDNGSGLLNMVQIAKGMNLEERHMRDLAARVNLVKTPOHVKI 695  
 868 TDFGLARLLDIDETEHADGKVPFKMALESILRRRFTHSDVMSYGVTVWELMTFGAK 927  
 696 TDFGLAKOLGADEKEHAEGKVPFKMALESILHRIYTHSDVMSYGVTVWELMTFGSK 755  
 928 PYDGIPIVAREIPDLLEKGERLPOPICTIVYIMVYCMIMIDSECRPFELVSEFSMAR 987  
 756 PYDGIPIVAREIPDLLEKGERLPOPICTIVYIMVYCMIMIDSECRPFELVSEFSMAR 815  
 988 DPQFVAVIO-NEDLGASPLDSTFYVSLLEDMDMDLVDAEVLVPOQGFPCDPDPAAG 1046  
 816 DPFVAVIOGDERMHLPSPTDSKFYRLTLEEDMEDIVDAEVLVPHQGF----- 866

QY 1047 GWHHRHSSSTRSGGDLTLGLEPSEEARPSPLAPSEGAGSDVFDGLGGAAGKIOS 1106  
 DB 867 -----NSPST-----SRTPLSLLSATSN-----NSATKICDR 894  
 QY 1107 LPTHPDPLQRYSEDPYVPLSETPGYVAPLTCQPEYVNOVDVROPSPREGLPAA 1166  
 DB 895 NGGH-----PYREDGFL-----PAPEYVNO-LMPKPSYAVYONQIY 930  
 QY 1167 RPAGAT-LEBAKTLSPKNGVYKDVAFGAGAVENPEYL 1203  
 DB 931 NYSILTAISKLPMSDRYQN-----SSTAVDNEEYL 961

RESULT 14  
 Q85468 PRELIMINARY; PRT; 545 AA.  
 AC Q85468;  
 DT 01-NOV-1996 (TEMBLrel. 01, Created)  
 DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)  
 DE 01-MAR-2002 (TEMBLrel. 20, Last annotation update)  
 DE Avian erythroblastosis virus (T834) v-erbB gene.  
 OS Avian erythroblastosis virus.  
 OC Viruses; Retroviridae; Avian type C retroviruses.  
 OX NCBI\_Taxid=11861;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=88217326; PubMed=2897102;  
 RA Scetling P., Vennstrom B., Jansen M., Graf T., Beug H., Hayman M.J.;  
 RT "Common site of mutation in the erbB gene of avian erythroblastosis  
 RT virus mutants that are temperature sensitive for transformation.";  
 RL Oncogene Res. 1:265-278(1987).  
 DR EMBL: X06943; CA30024.1; -.  
 DR HSSP: P11362; 1FGK.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00069; pkinase.1.  
 DR ProDom: PD000001; Euk\_pkinase.1.  
 DR SMART: SM00219; TYK1.1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP.1.  
 DR PROSITE: PS00011; PROTEIN\_KINASE\_DOM.1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR.1.  
 DR ATP-binding; Tyrosine-protein kinase.  
 KW Tyrosine-protein kinase.  
 SQ SEQUENCE 545 AA; 60899 MW; 140DCE8CA0F8AF CR664;

Query Match 22.9%; Score 1567; DB 15; Length 545;  
 Best Local Similarity 52.7%; Pred. No. 2e-110;  
 Matches 334; Conservative 73; Mismatches 129; Indels 98; Gaps 16;

578 GPEADQVCAHYKDPFCVACRCPSPGLSYMPIKFPDEGACOPCPINCHSCVDL 637  
 1 GP--DHCKCAHFIDGPHCVACPAVIGENDTL-VMKYADANAVCQLCHPCTRGCKGP 57  
 638 DDKGCPABQASPLTISVAVV-GILLVVLGVVFGILIKRQOKIRKXTMRRLLOETEL 696  
 58 GLEGP---NGSKTPSIAGVGGLLCLLVVVGIGLYLRR-HIVKRTLRLLQEREL 113  
 697 VEPLTPSGAMNQAMRILKEFNFTVSFMLRVKVSASHLETYKGIWIDGENVKI PV 756  
 114 VEPLTPSGEAPNQAHRLIKE-----TEFKYVVLGFGAFGLTYKGLIPEGKVTIPV 167  
 757 AIKVLRENTSPKANKELIDEAYVAVGSPYVSLGLICLTSTVOLVQLMPYGCCLDHY 816  
 168 AIKELRENTSPKANKELIDEAYVAVSDNPHVCGLLGICLTSTVOLVQLMPYGCCLDXY 227  
 228 REHNDNGSGLLNMVQIAKGMNLEERHMRDLAARVNLVKTPODVKITDGLAKOL 287  
 817 RENGRGLASQDLNMCQIAKGSYLEDVRLVHRDLAARVNLVKSPPHVKITDGLARLL 876  
 877 DIDETEHADGKVPFKMALESILRRRFTHSDVMSYGVTVWELMTFGAPYDGPARE 936  
 288 GADKEVHAEGKVPFKMALESILHRIYTHSDVMSYGVTVWELMTFGSKFYDGPARE 347

QY 937 IPDLKGERLPQPICTIDVYIMVCMIDSECRPRELVESEFARMADPQRFVYIQ 996  
 DB 348 ISSVLEKGERLPQPICTIDVYIMVCMIDSECRPRELVESEFARMADPQRFVYIQ 407  
 QY 997 NEULGPASPLDSTFFYSLEDDDMGDLVDAEELVYVQGFCCDPAPGAGGVHHRHS 1055  
 DB 408 GDERMHLPSPTSDSFYRTLMEEDEMDIVDAEVLPHQGF-----NS 451  
 QY 1056 SSTRSGGDLTLGEPSEEPASPL-----APEGAGSVDFGDLGMGAAGLQSLPTH 1110  
 DB 452 PST-----SRTPLLSSLSATSNMNTATNIDNG-----H 481  
 QY 1111 DPSPLOYSEDPVPLPSENDGYAPLTCSPQFPYVQPDVPPQSPSPREGPLPAAPAG 1170  
 DB 482 -----PYREDGFL-----PAPRYVQ--LMPKPSSTAMVQIQIYVYIS 517  
 QY 1171 AT-LEBAKTLSPKNGVYKDVAFAGAVENPEYL 1203  
 DB 518 LTAISKLPMSRYQ-----SHSTAVDNEPYL 544

## RESULT 15

Q9WVF5 PRELIMINARY; PRT; 655 AA.

AC Q9WVF5.  
 DT 01-NOV-1999 (TREMBlrel. 12, Created)  
 DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)  
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor (Epidermal growth factor receptor isoform 3).  
 GN EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RX [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=LIVER.  
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Schell C., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maibhe N.J.;  
 RT "Alternative transcripts from the Human and Mouse EGFR Genes Encode Carboxy-Terminal Truncated Receptors";  
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEVYAC;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Sinclair C.S., Pearse R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maibhe N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and mouse alternative Bgfr transcripts encoding truncated receptor isoforms";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=LIVER;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Aikawa K., Hara A., Fukumitsu Y., Komo H., Adachi J., Fukuda S., Aikawa K., Iwata M., Nishi K., Kiyosawa H., Kondo S., Yamana K.I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kato K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Flischnmann W., Gaasterland T., Gissi C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J., Schirai L.M., Staab F., Suzuki R., Tomita M., Wagner D., Washio T., Sasaki K., Okido T., Furuno M., Kono H., Baldarelli R., Barish G., Blake J., Botfield D., Boujunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,

RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F., Suzuki H., Togo-Oka K., Wang K.H., Weitz C., Whitteker C., Wilmink L., Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayaishiaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection";  
 RL Nature 409:685-690 (2001).  
 DR EMBL: AF124513; AAD44149.1; -  
 DR EMBL: AF275366; AAG28047.1; -  
 DR EMBL: AF275364; AAG28047.1; JOINED.  
 DR EMBL: AF275365; AAG28047.1; JOINED.  
 DR EMBL: AK004944; BAB23688.1; -  
 DR EMBL: AK004883; BAB23641.1; -  
 DR EMBL: AK004911; BAB23662.1; -  
 DR MGI:95294; Egfr.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR002174; Furin-like.  
 DR Pfam: PF01030; Furin-like.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR SMART: SM00261; FU; 3.  
 KW Receptor.  
 SQ SEQUENCE 655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match 22.0%; Score 1506.5; DB 11; Length 655;  
 Best Local Similarity 44.5%; Pred. No. 1e-105;  
 Matches 285; Conservative 97; Mismatches 233; Indels 25; Gaps 9;

QY 11 LIALALPQAA--STVCTGTDMKRLPASPETHLDMRLHYOGQVQGNLELYPTN 68  
 DB 14 LIALCAAGALAEKKVCGGTSNRLLQGTGFEDHLSLQRMNNEVGNLEILYVQGN 73  
 QY 69 ASLSPLQDIEVQGYVLAHNOVQVPLRLRIYVGTQLFEENYALAVLDNDPLNTTP 128  
 DB 74 YDLSFKTQIEVAGVYLAALNVERIPLENQIIRGNALYENTVALALISN----- 124  
 QY 129 VTGASPGSLREVLRLSLTEILKGVILQNPOLQYDTILMKDI----FKKNQALTLI 184  
 DB 125 -YGTNRTGRLPMLNLOGLLGAIRFSNPLTLCMDITLQREDIYQNVFMSMSDL--- 180  
 QY 185 DTRNSBACHPCSPMKGSRGSESSDQSLRTVYACGA-CRKGPLPTDCHQCAAG 243  
 DB 181 -QSHPSCKPCDPCSPGSCMGCGEENCQKTKIICAQQCSHRGSRSPDCCHQCAAG 239  
 QY 244 CGPKRSDCLACTHFNHSGICELHCPALVTYNTDFEEMPNREGYTGASCVTAPVY 303  
 DB 240 CTGPRSDCLVQKQDEATCKOTCPPLMLNPTTYQMDVNEGKTSRGATVKKCPRY 299  
 QY 304 LSTDVGSCTLVCPLNNOEVTAEQGTQCEKSKPCARVYGLGMQYIRANSKFIGITE 363  
 DB 300 VVTHDGSVYRACGPOYEV-EDGIRKCKKCGPCKVONGIGIEPK-DLISINATNIX 357  
 QY 364 -FAGCKTIGSLAFIPESFDGPAANTAPLEQIQLVETLEITGILYISAMPDLP 422  
 DB 358 HFKYCTALSGDHLIPVAFKGSFRTPLDRELEIKTYVEITGIFLLIQAMPNDWL 417  
 QY 423 SVFONLQVIRGRILHNGAVSLTLOGISLWLSRLSRLSGSLALIHNTHLCFHTVPM 482  
 DB 418 HAFENDEIRGRKTKHQGFSLAVLNLNITSGLRLKLSISDQVLIISGRNLCYANTINM 477  
 QY 483 DDLFRNPHQALHTNRPDEBCVGEGLACHOLCANHCWGPPTQVNCSCGLRQCEVE 542  
 DB 478 KLLFGPNQTKIKMNRKCKCKAVNHCNPLCSSEGGCGEPFRCVSCVNSRRECEV 537  
 QY 543 ECRVQLGPREVYVNRHCLPCGPCQONGSTGCPPEADQVACAHKDPFCVARGPS 602  
 DB 538 KNITLGEERREVENSECIQHFELPLQAMNITLGRGPDNLCIQCAHITDHPKVTCPA 597  
 QY 603 GVKPDLSTYPLWKPDEEGACQPCPINTGSCVDLDDKGC 642  
 DB 598 GINGENNTL-VKVKADANNVCHLCHANCTYGCAGGGLQGC 636

Search completed: July 22, 2003, 09:24.40  
 Job time : 53.0485 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:41:54 ; Search time 37.1104 Seconds  
(without alignments)  
4527.811 Million cell updates/sec

Title: SEQ4-695-709-14  
Perfect score: 6853  
Sequence: 1 MELALCRWGLLALLPPGA.....TRKGPTRAPNPEYLGLDVPV 1261

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues  
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: A\_Geneseq.101002.\*
- 2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*
- 3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*
- 4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*
- 5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*
- 6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*
- 7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*
- 8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*
- 9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*
- 10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*
- 11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*
- 12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*
- 13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*
- 14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*
- 15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*
- 16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*
- 17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*
- 18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*
- 19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*
- 20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*
- 21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*
- 23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6647	97.0	1255	21	AAV92620
2	6647	97.0	1255	22	AAE12130
3	6647	97.0	1255	22	AAE60167
4	6647	97.0	1255	23	AAU74545
5	6641	96.9	1255	17	AAW01111
6	6641	96.9	1255	21	AAW92406
7	6641	96.9	1255	21	AAAB2198
8	6641	96.9	1255	21	AAAB4780
9	6641	96.9	1255	22	AAAB5458
10	6641	96.9	1255	22	AAAB8267

11	6641	96.9	1255	23	AAE24067	Human Her-2 protei
12	6641	96.9	1255	23	AAE20479	Human Her-2/neu pr
13	6641	96.9	1255	23	AAE51143	Human Her-2/neu on
14	6641	96.9	1255	23	AAU77114	Human Her-2/neu po
15	6598	96.3	1433	14	AAE39568	Sequence of c-erbB
16	6477	94.5	1223	14	AAU98923	Human breast cance
17	6324	92.3	1200	21	AAAB21208	Human Her-2/neu pr
18	5854.5	85.4	1256	21	AAAB21199	Rat Her-2/neu prot
19	5854.5	85.4	1256	23	AAAB51144	Rat Her-2/neu onco
20	5829.5	85.1	1256	21	AAAB21206	Mouse Her-2/neu pr
21	5829.5	85.1	1256	22	AAAG2860	Amino acid sequenc
22	5829.5	85.1	1256	23	AAAB51151	Mouse Her-2/neu on
23	4811	70.2	919	23	AAAB21203	Human Her-2/neu fu
24	4811	70.2	919	23	AAAB51148	Her-2/neu extracel
25	4061.5	59.3	920	23	AAAB51152	Mouse Her-2/neu ex
26	4061.5	59.3	926	23	AAAB51153	Mouse Her-2/neu ex
27	3695	53.9	712	21	AAAB21204	Human Her-2/neu fu
28	3695	53.9	712	23	AAAB51149	Her-2/neu extracel
29	3552	51.8	782	18	AAAB19764	Her-2/neu-GFP immuno
30	3550	51.8	653	21	AAAB21200	Human Her-2/neu on
31	3550	51.8	653	21	AAAB51145	Human ErbB2 oncopr
32	3512	51.2	645	22	AAAB60408	Human ErbB2 extrac
33	3512	51.2	645	22	AAAB51593	DC8ecFv-erbB2EC fu
34	3447	50.3	951	21	AAAY4993	Extracellular port
35	3344	48.8	624	11	AAAB08222	Human EGF receptor
36	3065	44.7	1210	21	AAAB19259	Amino acid sequenc
37	3065	44.7	1210	21	AAAY50616	Human EGF receptor
38	3065	44.7	1210	23	AAE23019	Human Her-1 protei
39	3065	44.7	1210	23	AAAB50768	Human epidermal gr
40	3063	44.7	1210	22	AAAB68420	Amino acid sequenc
41	3053.5	44.6	654	21	AAAB21205	Rat Her-2/neu prot
42	3053.5	44.6	654	21	AAAB51150	Rat Her-2/neu onco
43	3024	43.7	1210	23	ABP51768	Human epidermal gr
44	2997	43.7	583	23	AAE20483	Human protein for
45	2997	43.7	587	23	AAE20481	Human protein for

## ALIGNMENTS

RESULT 1	AAV92620 standard; Protein, 1255 AA.
ID	AAV92620
AC	AAV92620;
XX	10-AUG-2000 (first entry)
DT	Human heregulin 2 (Her2).
XX	Her2; vaccination; cytotoxic T-lymphocyte immunity;
XX	self-protein; cancer; breast cancer; prostate cancer;
KW	cell-associated peptide antigen; foreign epitope.
XX	
OS	Homo sapiens.
XX	
PH	Key
FT	Domain
FT	Location/Qualifiers
FT	1..173
FT	/label= "N-terminal"
FT	/note= "mature polypeptide"
FT	5..25
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	59..73
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	103..117
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	149..163
FT	/label= "insertion region"
FT	/note= "suitable for foreign epitope insertion"
FT	174..332
FT	Domain



FT /label= Cysteine\_rich\_domain  
 FT Region  
 FT 210..224  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 250..264  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 324..483  
 FT /label= Ligand\_binding\_domain  
 FT 325..339  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 369..383  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 465..479  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 484..623  
 FT /label= Cysteine\_rich\_domain  
 FT 579..593  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 624..654  
 FT /label= Transmembrane\_domain  
 FT 632..652  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 653..667  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 655..1010  
 FT /label= Tyrosine\_kinase\_domain  
 FT 661..675  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 695..709  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 710..730  
 FT /label= insertion region  
 FT /note= "suitable for foreign epitope insertion"  
 FT 1011..1235  
 FT /label= C-terminal\_domain  
 XX  
 PN W0200020027-A2.  
 XX  
 PD 13-APR-2000.  
 XX  
 PF 05-OCT-1999; 99WC-DK00525.  
 XX  
 PR 05-OCT-1998; 98DK-0001261.  
 XX 20-OCT-1998; 98US-0105011.  
 PA (MEBT-) M & E BIOTECH AS.  
 PI Seinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I,  
 XX Gautam A, Birk P, Karlsson G;  
 XX WPI: 2000-34997/30.  
 DR N-PSDB; AAA09455.  
 PT  
 XX  
 PS Claim 62; Page 193-198; 220pp; English.  
 CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of  
 CC Her2 can be used in the claimed method as an autovaccine to induce a CTL  
 CC response. Subdominant CTL epitopes, antibody binding regions and  
 CC cysteine residues involved in disulfide bonds are preserved in the  
 CC immunogenized forms. Regions suitable for the insertion of foreign T  
 CC helper epitopes were identified (see features table). The method

CC is used for inducing immune responses against weakly immunogenic  
 CC cell-associated peptide antigens (PA) such as those associated with  
 CC cancers (self-proteins), e.g. human prostate specific membrane antigen  
 CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).  
 CC The method comprises effecting simultaneous presentation by antigen  
 CC producing cells (APCs) of the animal's immune system of: (1) at least 1  
 CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1  
 CC B-cell group derived from the cell-associated PA; and (2) at least 1  
 CC first T helper cell group which is foreign to the animal. Analogues of  
 CC human PSM, human Her2 and human/murine FGF8b comprising a substantial  
 CC part of all known and predicted CTL and B-cell epitopes of the respective  
 CC PA and including at least one foreign T helper epitope are also claimed.  
 CC The method is used to treat prostate, prostate/breast or breast cancer  
 CC when the PA is human PSM, FGF8b and Her2, respectively.  
 XX  
 SQ Sequence 1255 AA;

Query Match 97.0%; Score 6647; DB 21; Length 1255;  
 Best Local Similarity 97.5%; Pred. No. 0;  
 Matches 1230; Conservative 4; Mismatches 20; Indels 8; Gaps 2;

QY 1 METALCRWGLLALLPPGAASVCTGTDMKRLPASPETHLDMRLHYGQCQVQNL 60  
 DB 1 METALCRWGLLALLPPGAASVCTGTDMKRLPASPETHLDMRLHYGQCQVQNL 60  
 QY 61 EITVPTNASLFLDIOEVQGVYLIANNVROVPLQRLRYRGTLPEFNVALVDNG 120  
 DB 61 EITVPTNASLFLDIOEVQGVYLIANNVROVPLQRLRYRGTLPEFNVALVDNG 120  
 QY 121 DPLNNTPTVGTASPGGLRELQRLSTELTKGVYIQRPOLCYOFTIMKQIFHNQOLA 180  
 DB 121 DPLNNTPTVGTASPGGLRELQRLSTELTKGVYIQRPOLCYOFTIMKQIFHNQOLA 180  
 QY 121 DPLNNTPTVGTASPGGLRELQRLSTELTKGVYIQRPOLCYOFTIMKQIFHNQOLA 180  
 DB 121 DPLNNTPTVGTASPGGLRELQRLSTELTKGVYIQRPOLCYOFTIMKQIFHNQOLA 180  
 QY 181 LTLIDTNSRACHPCSPCKSCRCWSESDQSLTRTVACAGCARCKGPPDCHQC 240  
 DB 181 LTLIDTNSRACHPCSPCKSCRCWSESDQSLTRTVACAGCARCKGPPDCHQC 240  
 QY 241 AAGCTGPRHSDCLALHNHSGICGLHPALVTYVTFEEMPNEGYTGASVTCAP 300  
 DB 241 AAGCTGPRHSDCLALHNHSGICGLHPALVTYVTFEEMPNEGYTGASVTCAP 300  
 QY 301 YNYLSTDVSGCTVLCPLANOEVTAEDGTQRCCKSPCARCYGLGMOYIRANSKF 360  
 DB 301 YNYLSTDVSGCTVLCPLANOEVTAEDGTQRCCKSPCARCYGLGMOYIRANSKF 360  
 QY 361 EIEPAGCKKIFGSLAFLESPFGDPASTAPLQEPQLOVFTLEITGYLIISAMP 420  
 DB 361 EIEPAGCKKIFGSLAFLESPFGDPASTAPLQEPQLOVFTLEITGYLIISAMP 420  
 QY 421 DLSVFQNTQVIRGRILHNGAVSLTQGGISWLGRLSRLREISGLALIHNTLHCFV 480  
 DB 421 DLSVFQNTQVIRGRILHNGAVSLTQGGISWLGRLSRLREISGLALIHNTLHCFV 480  
 QY 481 PMDQLFRPHQALLHTANRPEDEVCYEGSLAQCLCARHCHGPGPTQCVNSQFLRG 540  
 DB 481 PMDQLFRPHQALLHTANRPEDEVCYEGSLAQCLCARHCHGPGPTQCVNSQFLRG 540  
 QY 541 VEECVTLGLPREYVNAHCLPCHECQPONGSVTCFEPADQCAAHYDPPCVARC 600  
 DB 541 VEECVTLGLPREYVNAHCLPCHECQPONGSVTCFEPADQCAAHYDPPCVARC 600  
 QY 601 PEGVAPDLSYMPIMKFPDEEGACQPCPINTHSCVDLDDKGPABGRASPISIVAVG 660  
 DB 601 PEGVAPDLSYMPIMKFPDEEGACQPCPINTHSCVDLDDKGPABGRASPISIVAVG 660  
 QY 661 ILVVAVLVVREGILIKRQOKIKRYTMRLLQETLVPLTPSGAMPQAOQRLIKERE 720  
 DB 661 ILVVAVLVVREGILIKRQOKIKRYTMRLLQETLVPLTPSGAMPQAOQRLIKERE 720  
 QY 721 FTVSFWLVPKXSAS-HLETYVKGIMIDGENVKIPVAKYLRNTSPKANKELIDEAYV 779  
 DB 720 -----LKKVYLVGAGGATYVKGIMIDGENVKIPVAKYLRNTSPKANKELIDEAYV 773

```

Qy 780 MAGVSPYVSLIGLITSTVQLVTLQMLMYGCLLDHVRNRLGSLQDILLNCMOIAGKM 839
Db 774 MAGVSPYVSLIGLITSTVQLVTLQMLMYGCLLDHVRNRLGSLQDILLNCMOIAGKM 833
Qy 840 SYLEDVRLVHRDLAANVLYKSPNHVKTDFGLARLLDIDETEVHADGGKVIKMMALLES 899
Db 834 SYLEDVRLVHRDLAANVLYKSPNHVKTDFGLARLLDIDETEVHADGGKVIKMMALLES 893
Qy 900 IIRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLKEGRLPQPICTIDVYM 959
Db 894 IIRRRFTQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLKEGRLPQPICTIDVYM 953
Qy 960 IMVKCWMIDSECRPRFRELVSERSRKARDPQRFVVIQNDLGPASPLDSTFRSLLEDD 1019
Db 954 IMVKCWMIDSECRPRFRELVSERSRKARDPQRFVVIQNDLGPASPLDSTFRSLLEDD 1013
Qy 1020 MGDLVDAEELVLPQCGFPCPDPAFGAGVWHHRHSSSTRSGGDLTLGLFSESEAPRS 1079
Db 1014 MGDLVDAEELVLPQCGFPCPDPAFGAGVWHHRHSSSTRSGGDLTLGLFSESEAPRS 1073
Qy 1080 PLAPSEAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYAAPLTC 1139
Db 1074 PLAPSEAGSDVFDGLGMAAKGLQSLPTHDPSPLOQRYSEDPVPLPSETDGYAAPLTC 1133
Qy 1140 SPOPEYVNOVDYRPPSPREBGLPARAPAGATLEBAKTLSGKGVVWDVAFEGAVEN 1199
Db 1134 SPOPEYVNOVDYRPPSPREBGLPARAPAGATLEBAKTLSGKGVVWDVAFEGAVEN 1193
Qy 1200 PEYLTPQGAAPQHPHPAPFSAFDMLYWDDQPEPERGAPSTFKGTPTAENPEVILGDV 1259
Db 1194 PEYLTPQGAAPQHPHPAPFSAFDMLYWDDQPEPERGAPSTFKGTPTAENPEVILGDV 1253
Qy 1260 PV 1261
Db 1254 PV 1255

RESULT 2
ID AAE12130 standard; Protein; 1255 AA.
AAE12130;
AC AAE12130;
DT 18-DEC-2001 (first entry)
DE Human tyrosine kinase-type receptor, HER-2.
KW Therapeutic compound; major histocompatibility complex; vaccine;
KW antigenic peptide; MHC; immunoregulatory; immune response; HER-2;
KW adoptive immunotherapy; anti-cancer; breast cancer antigen; APC;
KW antigen presenting cell; human; tyrosine kinase-type receptor.
OS Homo sapiens.
XX
XX
XX Key Location/Qualifiers
XX Region 774..782
XX /note="Antigenic epitope"
XX
XX WO200168677-A2.
XX
XX 20-SEP-2001.
XX
XX 16-MAR-2001; 2001WO-US40328.
XX PF 16-MAR-2000; 2000US-0527487.
XX PR 16-MAR-2000; 2000US-0527487.
XX
XX PA (GENZ ) GENZYME CORP.
XX PI Nicolette CA;
XX
XX WPI; 2001-616284/71.
XX DR N-PsDB; AAD19731.
XX

```

PT Novel synthetic therapeutic compound for inducing immune response and  
 PT for use in adoptive immunotherapy, has enhanced binding to major  
 PT histocompatibility molecules and enhanced immunoregulatory properties  
 PT

PS Claim 4; Page 63-67; 69pp; English.

CC The invention relates to synthetic therapeutic compounds (antigenic  
 CC peptides) with enhanced binding to major histocompatibility complex  
 CC (MHC) molecules and enhanced immunoregulatory properties relative  
 CC to their natural counterparts. Compounds of the invention are useful  
 CC for inducing an immune response in a subject and for use in adoptive  
 CC immunotherapy. They are useful as components of anti-cancer vaccines  
 CC and to expand immune effector cells that are specific for cancers  
 CC characterised by expression of the breast cancer antigen, HER-2.  
 CC Polynucleotides that encode peptides of the invention are useful as  
 CC hybridisation probes and as primers for the detection of genes of  
 CC transcripts that are expressed in antigen presenting cells (APCs), to  
 CC confirm transduction of polynucleotides into host cells. The present  
 CC sequence is human tyrosine kinase-type receptor, HER-2. Compounds  
 CC of the invention are designed based on the HER-2 antigenic peptide  
 CC (774-782).

CC Sequence 1255 AA;

Query Match 97.0%; Score 6647; DB 22; Length 1255;  
 Best Local Similarity 97.5%; Pred. No. 0;  
 Matches 1230; Conservative 4; Mismatches 20; Indels 8; Gaps 2;

```

Qy 1 MELALCRWGLLLLPAGASTQVCGTGMKRLPASPETHLMDLHLYGCGVQGNL 60
Db 1 MELALCRWGLLLLPAGASTQVCGTGMKRLPASPETHLMDLHLYGCGVQGNL 60
Qy 61 ELTYLPTNASLFLQDIQEVQGYVLIANQVQYPLQRLIRVGTQLFEDNYALAVDNG 120
Db 61 ELTYLPTNASLFLQDIQEVQGYVLIANQVQYPLQRLIRVGTQLFEDNYALAVDNG 120
Qy 121 DPLNNTTPVYGASGGRRLQRLSTLILKGVLIQNPOLCIYDITLWKDIFHKNNOLA 180
Db 121 DPLNNTTPVYGASGGRRLQRLSTLILKGVLIQNPOLCIYDITLWKDIFHKNNOLA 180
Qy 121 DPLNNTTPVYGASGGRRLQRLSTLILKGVLIQNPOLCIYDITLWKDIFHKNNOLA 180
Db 121 DPLNNTTPVYGASGGRRLQRLSTLILKGVLIQNPOLCIYDITLWKDIFHKNNOLA 180
Qy 181 LTLIDTRSRACHPCSPCMKSGRCWGSSESDQSLTRTVACAGCARCKGPLPTCCHEOC 240
Db 181 LTLIDTRSRACHPCSPCMKSGRCWGSSESDQSLTRTVACAGCARCKGPLPTCCHEOC 240
Qy 241 AAGCTGPKHSDCLACLPHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
Db 241 AAGCTGPKHSDCLACLPHNSGICELHCPALVTYNTDTFESMPNDEGRYTFGASCVTACP 300
Qy 301 YNVLSTVGSCTLVCPHNOEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
Db 301 YNVLSTVGSCTLVCPHNOEVTAEDGTQRCCKSKPCARVCYGLGMEHLREVAVTSAN 360
Qy 361 ELBPAGCKKIFGSLAFIPESFDDGPASNTAPLOEQVFEETLEITGYLYISAMPDPLP 420
Db 361 ELBPAGCKKIFGSLAFIPESFDDGPASNTAPLOEQVFEETLEITGYLYISAMPDPLP 420
Qy 421 DLSVFQNLQVIRGRIILNNGAYSLTLQGLGISWGLRLSRLGSLALIHNTHLCEVHTV 480
Db 421 DLSVFQNLQVIRGRIILNNGAYSLTLQGLGISWGLRLSRLGSLALIHNTHLCEVHTV 480
Qy 481 PMDQLFANPQALHTANRPEDECVGEGLAHQULCARGHGMSGPPTQCVNCSQFLRQEC 540
Db 481 PMDQLFANPQALHTANRPEDECVGEGLAHQULCARGHGMSGPPTQCVNCSQFLRQEC 540
Qy 541 VEEGRVLOGLPREYVNAHCLPCHPECOPNGSVTCGPEADOCVACAHYKDPFCVACR 600
Db 541 VEEGRVLOGLPREYVNAHCLPCHPECOPNGSVTCGPEADOCVACAHYKDPFCVACR 600
Qy 601 PSGVYKPLSYMPIMKPEDEGACOPCINCTHS CVLDDGCAEAGASPLTISAVNG 660
Db 601 PSGVYKPLSYMPIMKPEDEGACOPCINCTHS CVLDDGCAEAGASPLTISAVNG 660

```

QY 661 ILLVVLGVVFGIILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNOAMRIKEFN 720  
 DB 661 ILLVVLGVVFGIILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNOAMRIKEFE- 719  
 QY 721 FTVSFVLVVKVASAS-HLETYKGIWIPDGENVKIPVAIKVLRENTSPPANKELIDEAVY 779  
 DB 720 -----LRKVKVIGSGAFGTIVYKGIWIPDGENVKIPVAIKVLRENTSPPANKELIDEAVY 773  
 QY 780 MAGVGSPPYVSRLLGICITSTVQLVLTQMPYGLLDHYRENRKGLSGQDLINMCMQIAKGM 839  
 DB 774 MAGVGSPPYVSRLLGICITSTVQLVLTQMPYGLLDHYRENRKGLSGQDLINMCMQIAKGM 833  
 QY 840 SYLEDVVLVVRDLAARNVVKSPNHVKITDFGLARLLIDETEHYHAGGKVPKIMWALS 899  
 DB 834 SYLEDVVLVVRDLAARNVVKSPNHVKITDFGLARLLIDETEHYHAGGKVPKIMWALS 893  
 QY 900 ILRRFTHSGDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 959  
 DB 894 ILRRFTHSGDWSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953  
 QY 960 IMVKCMWIDSECRPRELVSSESRMARDPQRFVIONEDLGASPLDSTFYRSLLEDD 1019  
 DB 954 IMVKCMWIDSECRPRELVSSESRMARDPQRFVIONEDLGASPLDSTFYRSLLEDD 1013  
 QY 1020 MGDVLDAEYLVPQCGFFCPDPAAGAGVNHHRSSSTRSGGDLTLGLEPSEEPARS 1079  
 DB 1014 MGDVLDAEYLVPQCGFFCPDPAAGAGVNHHRSSSTRSGGDLTLGLEPSEEPARS 1073  
 QY 1080 PLAPSEGASDVFDGDLGMAAAGLQSLPTHDSPLQRYSEDPVLPSTDTGYVAPLTC 1139  
 DB 1074 PLAPSEGASDVFDGDLGMAAAGLQSLPTHDSPLQRYSEDPVLPSTDTGYVAPLTC 1133  
 QY 1140 SPOPEYVNOVDVFPQPSPREGLPAAPAGATLEBAKTISPKNQVWQVFAFGAVEN 1199  
 DB 1134 SPOPEYVNOVDVFPQPSPREGLPAAPAGATLEBAKTISPKNQVWQVFAFGAVEN 1193  
 QY 1200 PEYVLTQGAAPQHPHPAPSPAFDNLVYWDQPPERGAPSTFFKGTPTANPEYVGLDV 1259  
 DB 1194 PEYVLTQGAAPQHPHPAPSPAFDNLVYWDQPPERGAPSTFFKGTPTANPEYVGLDV 1253  
 QY 1260 PV 1261  
 DB 1254 PV 1255

RESULT 3  
 ID AAB60167 standard; Protein: 1255 AA.  
 AC AAB60167;  
 DT 03-APR-2001 (first entry)  
 DE HHR2 transgene plasmid construct encoded protein.  
 KM Human, HHR2; ErbB2 receptor; p185neu; maytansinoid conjugate; cancer;  
 OS antibody.  
 OS Homo sapiens.  
 OS Synthetic.  
 PN M0200100244-A2.  
 PD 04-JAN-2001.  
 PF 23-JUN-2000; 2000WO-US17229.  
 PR 25-JUN-1999; 99US-0141316.  
 PR 16-MAR-2000; 2000US-0189844.  
 PA (GETH) GENENTECH INC.  
 PI Erickson S, Schwall R;

XX MPI: 2001-061962/07.  
 DR N-ESDB: AAF24297.  
 XX  
 PT Treating tumors, particularly breast cancers, which overexpress an ErbB  
 PT receptor and does not respond to an anti-ErbB antibody, comprises  
 PT conjugating the antibody to a maytansinoid -  
 XX  
 PS Example 3; Fig 4; 92pp; English.  
 XX  
 CC The present invention provides a method of treating cancer by  
 CC administering a conjugate of anti-ErbB antibody with a maytansinoid. In  
 CC particular, the antibody is directed against ErbB2 (also known as HER2  
 CC and p185neu). The method is particularly useful in the treatment of  
 CC breast, ovarian, stomach, endometrial, salivary gland, lung, kidney,  
 CC colon, colorectal, thyroid, pancreatic, prostate and bladder cancers.  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 97.0%; Score 6647; DB 22; Length 1255;  
 Best Local Similarity 97.5%; Pred. No. 0;  
 Matches 1230; Conservative 4; Mismatches 20; Indels 8; Gaps 2;  
 QY 1 METALCRWGLLALILPPGAASTVCTGTDMKRLPASPEHLMRLHYGCGQVYQNL 60  
 DB 1 METALCRWGLLALILPPGAASTVCTGTDMKRLPASPEHLMRLHYGCGQVYQNL 60  
 QY 61 ELTYLPTNASSFLDIDIEVGVYLIANOVQVPLQRLRYRGTOLEFEDNYALAVLDNG 120  
 DB 61 ELTYLPTNASSFLDIDIEVGVYLIANOVQVPLQRLRYRGTOLEFEDNYALAVLDNG 120  
 QY 121 DELNNTPTVTGASPGGLRELOLRSTELIKKGVLIQRNPQLCYODTILMKDIFHNQOLA 180  
 DB 121 DELNNTPTVTGASPGGLRELOLRSTELIKKGVLIQRNPQLCYODTILMKDIFHNQOLA 180  
 QY 181 LTLIDTNSRACHPCSPCKSGRCWSSSEDCOSLTRVCAAGGACRCKGPLEPTDCHEQC 240  
 DB 181 LTLIDTNSRACHPCSPCKSGRCWSSSEDCOSLTRVCAAGGACRCKGPLEPTDCHEQC 240  
 QY 241 AAGCTGPRHSOCIALFHNHSGICEHCPALVTYNTDFEEMPNPEGRTYGASCVTACP 300  
 DB 241 AAGCTGPRHSOCIALFHNHSGICEHCPALVTYNTDFEEMPNPEGRTYGASCVTACP 300  
 QY 301 YNYLSTDVGSCTLVCPRLNQEVTADGTQRCCKSKPCARVCYGIQMOYIKANSKFIGIT 360  
 DB 301 YNYLSTDVGSCTLVCPRLNQEVTADGTQRCCKSKPCARVCYGIQMOYIKANSKFIGIT 360  
 QY 361 ELIEPAGCKKIFGSLAFLESPFDGDPASTAFLQBPOLQVETLEITGYLYISAMPDLP 420  
 DB 361 IOEFAGCKKIFGSLAFLESPFDGDPASTAFLQBPOLQVETLEITGYLYISAMPDLP 420  
 QY 421 DLSTVQNLQVIRGRILHNGAYSLTQGGISMLGLRSRLREGSGLATLHNHTHLCFVHTV 480  
 DB 421 DLSTVQNLQVIRGRILHNGAYSLTQGGISMLGLRSRLREGSGLATLHNHTHLCFVHTV 480  
 QY 481 PWDQLFRNPHQALLHTANRPEDECVGBGLACHQLCARHCGMPGTQCVNSQPLRGQEC 540  
 DB 481 PWDQLFRNPHQALLHTANRPEDECVGBGLACHQLCARHCGMPGTQCVNSQPLRGQEC 540  
 QY 541 VEEGRVLOGLPREYNARHCLPCHPECOPONGSVTCFGEADQCAVCHYDPPFCVARG 600  
 DB 541 VEEGRVLOGLPREYNARHCLPCHPECOPONGSVTCFGEADQCAVCHYDPPFCVARG 600  
 QY 601 PSGVAPDLSYMPIWKFPEDEGACPCPINTCHSCVDLDKGCAPAEORASPLTISAVAVG 660  
 DB 601 PSGVAPDLSYMPIWKFPEDEGACPCPINTCHSCVDLDKGCAPAEORASPLTISAVAVG 660  
 QY 661 ILLVVLGVVFGIILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNOAMRIKEFN 720  
 DB 661 ILLVVLGVVFGIILIKRROOKIRKYMRLLOETELVEPLTPSGAMPNOAMRIKEFE- 719  
 QY 721 FTVSFVLVVKVASAS-HLETYKGIWIPDGENVKIPVAIKVLRENTSPPANKELIDEAVY 779

Db 720 -----LRKXVLSGSAFGTVYKGIWIPDGENKIPVIAIVLENTSPKANKETLBEAYV 773  
 Qy 780 MAGVSPYVSRLIGICTSTVQVLTQMPYGCILLDHVRENKRGSLDLMWCMQIAKGM 839  
 Db 774 MAGVSPYVSRLIGICTSTVQVLTQMPYGCILLDHVRENKRGSLDLMWCMQIAKGM 833  
 Qy 840 SYLEDVLRHDLAARVAVKSPHVKITDIFGLARLLDIBETVHADGKVPKMMALBS 899  
 Db 834 SYLEDVLRHDLAARVAVKSPHVKITDIFGLARLLDIBETVHADGKVPKMMALBS 893  
 Qy 900 ILRRRFTHSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 959  
 Db 894 ILRRRFTHSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPICTIDVYM 953  
 Qy 960 IMKCMWIDSECRREFELVSEFSRMARDQRFVYVIONEDLGPASPLDSTFFYSLLIEDD 1019  
 Db 954 IMKCMWIDSECRREFELVSEFSRMARDQRFVYVIONEDLGPASPLDSTFFYSLLIEDD 1013  
 Qy 1020 MGDVLDVDAEYLVPQGFPCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRS 1079  
 Db 1014 MGDVLDVDAEYLVPQGFPCPDPAAGAGMWHHRSSSTRSGGDLTLGLEPSEEEAPRS 1073  
 Qy 1080 PLAPSEGASDVFDGDIKMGAAKGLSLPHTDPSLQRYSEDPVPLPSTDGYVAPLTC 1139  
 Db 1074 PLAPSEGASDVFDGDIKMGAAKGLSLPHTDPSLQRYSEDPVPLPSTDGYVAPLTC 1133  
 Qy 1140 SPOEYVNOPDVRPQPSPREGPPAPAPAGATLEAKTSLSPKNGVWCVAFGAVEN 1199  
 Db 1134 SPOEYVNOPDVRPQPSPREGPPAPAPAGATLEAKTSLSPKNGVWCVAFGAVEN 1193  
 Qy 1200 PEYLTPQGAAPQHPHPAPSPAFDNLVYWDQDPPEGAPPTFKGTPTAENPEYGLDV 1259  
 Db 1194 PEYLTPQGAAPQHPHPAPSPAFDNLVYWDQDPPEGAPPTFKGTPTAENPEYGLDV 1253  
 Qy 1260 PV 1261  
 Db 1254 PV 1255  
 Db  
 RESULT 4  
 ID AAU74545 standard; Protein: 1255 AA.  
 AC AAU74545;  
 DT 23-APR-2002 (first entry)  
 DE Human HER2 (ErbB2) polypeptide.  
 XX Human; HER2; ErbB; epidermal growth factor receptor; receptor;  
 KW anti-ErbB antibody-maytansinoid conjugate; cancer; tumour; breast; ovary;  
 KW stomach; endometrium; salivary gland; lung; kidney; colon; colorectum;  
 KW thyroid; pancreas; prostate; bladder; ErbB2; neuronal disorder;  
 KW glial disorder; astrocytic disorder; hypochalamic disorder;  
 KW glandular disorder; macrophagal disorder; epithelial disorder;  
 KW stromal disorder; blastocoelec disorder; inflammatory disorder;  
 KW angiogenic disorder; immunological disorder.  
 XX Homo sapiens.  
 OS  
 PN US2002001587-A1.  
 PD 03-JAN-2002.  
 XX  
 PF 16-MAR-2001; 2001US-0811123.  
 PR 16-MAR-2000; 2000US-189844P.  
 PR 05-OCT-2000; 2000US-238327P.  
 XX  
 PA (ERIC/) ERICKSON S.  
 PA (SCHW/) SCHWALL R.  
 PA (SLIW/) SLIWKOWSKI M.

PI Erickson S, Schwall R, Slivkowski M;  
 XX WPI: 2002-163686/21.  
 DR N-FSD; ABK14058.  
 XX  
 PT Treating tumour characterised by overexpression of epidermal growth  
 factor receptor, ErbB or cancer in mammal, comprises administering  
 PT anti-ErbB antibody-maytansinoid conjugate to the mammal  
 XX  
 PS Example 3; Fig 7; 93pp; English.  
 CC The invention relates to treating a tumour in a mammal, where the tumour  
 CC is characterised by the overexpression of an epidermal growth factor  
 CC receptor (ErbB) and does not respond or responds poorly, to treatment  
 CC with an anti-ErbB antibody, comprising administering to the mammal an  
 CC anti-ErbB antibody-maytansinoid conjugate. The method is useful for  
 CC treating cancer or tumours of the breast, ovary, stomach, endometrium,  
 CC salivary gland, lung, kidney, colon, colorectum, thyroid, pancreas,  
 CC prostate and bladder, preferably breast cancer. The breast cancer is a  
 CC metastatic breast cancer or an aggressive form of metastatic breast  
 CC cancer which overexpresses ErbB2. The method is also useful for treating  
 CC neuronal, glial, astrocytic, hypochalamic, glandular, macrophagal,  
 CC epithelial, stromal, blastocoelec, inflammatory, angiogenic and  
 CC immunological disorders. This sequence represents the human HER2 (ErbB2)  
 CC polypeptide of the invention.  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 97.0%; Score 6647; DB 23; Length 1255;  
 Best Local Similarity 97.5%; Pred. No. 0;  
 Matches 1230; Conservative 4; Mismatches 20; Indels 8; Gaps 2;  
 Qy 1 MELALCRWGLLALLPPGAASYCTGTDMKLRIPASBETHLDMRLHYGCGVYQNL 60  
 Db 1 MELALCRWGLLALLPPGAASYCTGTDMKLRIPASBETHLDMRLHYGCGVYQNL 60  
 Qy 61 ETTYPTNASSIFLDIOEVQGYVIAHNOYVPLQRLRYRGTLFEDNYALAVLNG 120  
 Db 61 ETTYPTNASSIFLDIOEVQGYVIAHNOYVPLQRLRYRGTLFEDNYALAVLNG 120  
 Qy 61 ETTYPTNASSIFLDIOEVQGYVIAHNOYVPLQRLRYRGTLFEDNYALAVLNG 120  
 Db 61 ETTYPTNASSIFLDIOEVQGYVIAHNOYVPLQRLRYRGTLFEDNYALAVLNG 120  
 Qy 121 DELNNTTPTVGTASPGGLRELOLSLTELKGVLIQRNPOLCYOQTLIMKDIFFHKNOLA 180  
 Db 121 DELNNTTPTVGTASPGGLRELOLSLTELKGVLIQRNPOLCYOQTLIMKDIFFHKNOLA 180  
 Qy 121 DELNNTTPTVGTASPGGLRELOLSLTELKGVLIQRNPOLCYOQTLIMKDIFFHKNOLA 180  
 Db 121 DELNNTTPTVGTASPGGLRELOLSLTELKGVLIQRNPOLCYOQTLIMKDIFFHKNOLA 180  
 Qy 181 LTLIDTNSRACHPSPCKSGSRCKGSESSDQSLTRVYAGGACRGPLPTDCCHQC 240  
 Db 181 LTLIDTNSRACHPSPCKSGSRCKGSESSDQSLTRVYAGGACRGPLPTDCCHQC 240  
 Qy 181 LTLIDTNSRACHPSPCKSGSRCKGSESSDQSLTRVYAGGACRGPLPTDCCHQC 240  
 Db 181 LTLIDTNSRACHPSPCKSGSRCKGSESSDQSLTRVYAGGACRGPLPTDCCHQC 240  
 Qy 241 AAGCTGPRGSDCLAHFNHSGICELHCPALVTYNTDFEEMPNREGRYTGASCVTACP 300  
 Db 241 AAGCTGPRGSDCLAHFNHSGICELHCPALVTYNTDFEEMPNREGRYTGASCVTACP 300  
 Qy 241 AAGCTGPRGSDCLAHFNHSGICELHCPALVTYNTDFEEMPNREGRYTGASCVTACP 300  
 Db 241 AAGCTGPRGSDCLAHFNHSGICELHCPALVTYNTDFEEMPNREGRYTGASCVTACP 300  
 Qy 301 YNYLSTDVSGCTVCPPLHNOEVTADGTORCEKSKPCARVYCYGLMGIYKANSKFIGIT 360  
 Db 301 YNYLSTDVSGCTVCPPLHNOEVTADGTORCEKSKPCARVYCYGLMGIYKANSKFIGIT 360  
 Qy 301 YNYLSTDVSGCTVCPPLHNOEVTADGTORCEKSKPCARVYCYGLMGIYKANSKFIGIT 360  
 Db 301 YNYLSTDVSGCTVCPPLHNOEVTADGTORCEKSKPCARVYCYGLMGIYKANSKFIGIT 360  
 Qy 361 EUEFAGCKKITGSLAFIPSEFDGDPASNTALQSEOLQVETLEITGYLISAMPDLP 420  
 Db 361 EUEFAGCKKITGSLAFIPSEFDGDPASNTALQSEOLQVETLEITGYLISAMPDLP 420  
 Qy 361 EUEFAGCKKITGSLAFIPSEFDGDPASNTALQSEOLQVETLEITGYLISAMPDLP 420  
 Db 361 EUEFAGCKKITGSLAFIPSEFDGDPASNTALQSEOLQVETLEITGYLISAMPDLP 420  
 Qy 421 DLSVFONLQVIRGRILHNGAYSLTLQIGISWLGRLSRLREGSGALIHNTHTCFYATV 480  
 Db 421 DLSVFONLQVIRGRILHNGAYSLTLQIGISWLGRLSRLREGSGALIHNTHTCFYATV 480  
 Qy 421 DLSVFONLQVIRGRILHNGAYSLTLQIGISWLGRLSRLREGSGALIHNTHTCFYATV 480  
 Db 421 DLSVFONLQVIRGRILHNGAYSLTLQIGISWLGRLSRLREGSGALIHNTHTCFYATV 480  
 Qy 481 PMDOLFRPHALHTANRPEDCEVGSGLACHQICAGHCHGPGPTCCVNSQPLRQEC 540  
 Db 481 PMDOLFRPHALHTANRPEDCEVGSGLACHQICAGHCHGPGPTCCVNSQPLRQEC 540  
 Qy 481 PMDOLFRPHALHTANRPEDCEVGSGLACHQICAGHCHGPGPTCCVNSQPLRQEC 540  
 Db 481 PMDOLFRPHALHTANRPEDCEVGSGLACHQICAGHCHGPGPTCCVNSQPLRQEC 540  
 Qy 541 VEECRVLTGGLPREYVNAHCHLCPHPECPQNGSITCGPREADQVACAHYDPPFCVARC 600  
 Db 541 VEECRVLTGGLPREYVNAHCHLCPHPECPQNGSITCGPREADQVACAHYDPPFCVARC 600  
 Qy 541 VEECRVLTGGLPREYVNAHCHLCPHPECPQNGSITCGPREADQVACAHYDPPFCVARC 600  
 Db 541 VEECRVLTGGLPREYVNAHCHLCPHPECPQNGSITCGPREADQVACAHYDPPFCVARC 600  
 Qy 601 PSSGKPLDSTYMPKPFDEBGACQPCPINCCTHSCVDLDDKGCAPRAPSPLTISVAVG 660  
 Db 601 PSSGKPLDSTYMPKPFDEBGACQPCPINCCTHSCVDLDDKGCAPRAPSPLTISVAVG 660

```

Db      601 PSQVAPDLSTYMPKPFDEEGACQPCPINCSTHSCVDDLDKCCPAEQASPLTISVAVVG 660
Qy      661 ILLVVLGVVFGILIKRQOKIRKYTMRLQETELVEPLPSGAMNQAMRLKEFNN 720
Db      661 ILLVVLGVVFGILIKRQOKIRKYTMRLQETELVEPLPSGAMNQAMRLKEFNN 719
Qy      721 FTVSFWLRVPKVSAS-HLETVYKGIWIPDGENVKI PAIKYLRENTS PKANKELIDEAYV 779
Db      720 -----LRKVKVLSSGAFGVYKGIWIPDGENVKI PAIKYLRENTS PKANKELIDEAYV 773
Qy      780 MAGVSPVSRLLGLCLTSTVQLVTOAMPYGLLDHRENRGRSGODLLNMCQIAKGM 839
Db      774 MAGVSPVSRLLGLCLTSTVQLVTOAMPYGLLDHRENRGRSGODLLNMCQIAKGM 833
Qy      840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALES 899
Db      834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADGKVPKIMMALES 893
Qy      900 ILRRRFTHQSDVMSVGTVMELMTFGAKPYDGI PARSLPDLLEKGERLPPOPPICTIDVYM 959
Db      894 ILRRRFTHQSDVMSVGTVMELMTFGAKPYDGI PARSLPDLLEKGERLPPOPPICTIDVYM 953
Qy      960 IMVKWMIDSECRFRFELVSEFSRMAADPQRFVVIQNEIDGSPASPLDSTFYSLIEDDD 1019
Db      954 IMVKWMIDSECRFRFELVSEFSRMAADPQRFVVIQNEIDGSPASPLDSTFYSLIEDDD 1013
Qy      1020 MGDLVDAEYLVPOQGFCDPAPAGAGMHHRRSSSTSGGGLTLGLPSEEEAPRS 1079
Db      1014 MGDLVDAEYLVPOQGFCDPAPAGAGMHHRRSSSTSGGGLTLGLPSEEEAPRS 1073
Qy      1080 PLASEGAGSDVDFGDDGMGAAGLQSLPTHDPSPQLRYSEDDPVLPSESTDVYVAPLTC 1139
Db      1074 PLASEGAGSDVDFGDDGMGAAGLQSLPTHDPSPQLRYSEDDPVLPSESTDVYVAPLTC 1133
Qy      1140 SPQSEYVNOQDVRPQPSPREGLPAARPAATLERAKTILSPGNKGVVXOVFAFGAVEN 1199
Db      1134 SPQSEYVNOQDVRPQPSPREGLPAARPAATLERAKTILSPGNKGVVXOVFAFGAVEN 1193
Qy      1200 PEYLTPOGGAAPQHPAPSPAPDNLVYWDODPREGAPSTFKGTPTENBEYGLDV 1259
Db      1194 PEYLTPOGGAAPQHPAPSPAPDNLVYWDODPREGAPSTFKGTPTENBEYGLDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

```

## RESULT 5

```

AAW01111
ID      AAW01111 standard; Protein; 1255 AA.
AC      AAW01111;
XX
XX
DT      01-JAN-1997 (first entry)
XX
XX      HER-2/neu protein.
XX
XX      HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;
XX      breast cancer; ovary cancer; colon cancer; lung cancer;
XX      prostate cancer; immunisation; tumour; vaccine; vector.
XX
XX      Homo sapiens.
XX
XX      Key
XX      Domain
XX      Location/Qualifiers
XX      /label= Intracellular domain
XX      /note= "claimed domain, useful for immunisation"
XX
XX      W09630514-A1.
XX
XX      03-OCT-1996.
XX

```

```

PF      28-MAR-1996; 96MO-US01689.
XX
PR      31-MAR-1995; 95US-0414417.
XX
PA      (UNIT) UNIT WASHINGTON.
XX
PI      Cheever MA, Disis ML;
XX
DR      WPI; 1996-455361/45.
XX
DR      N-PSDB; AAT40739.
XX
PT      DNA encoding HER-2-neu poly-peptide(s) - used for prevention or
PT      treatment of malignancies with which the HER-2/neu oncogene is
PT      associated
XX
PS      Claim 2; Page 56-61; 71pp; English.
XX
CC      Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is
CC      the product of the HER-2/neu oncogene (see also AAT40739). The
CC      protein is over-expressed in various cancers, including breast,
CC      ovarian, colon, lung and prostate. The intracellular domain of the
CC      protein can be used to immunise an animal against a malignancy with
CC      which the oncogene is associated. The polypeptide can be produced
CC      in transformed host cells for use in immunisation. Alternatively,
CC      animal cells are transfected in vivo or ex vivo with a viral vector
CC      that directs expression of the polypeptide.
XX
SQ      Sequence 1255 AA;

```

Query Match 96.9%; Score 6641; DB 17; Length 1255;  
Best Local Similarity 97.3%; Pred. No. 0;  
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;

```

Qy      1 MEALALCWGLLALLPPGAASVQCTGTDMLRLPASPETHLMDLRHLYGSCQVVGNTL 60
Db      1 MEALALCWGLLALLPPGAASVQCTGTDMLRLPASPETHLMDLRHLYGSCQVVGNTL 60
Qy      61 ELTYLPTNASLFLDIOEVQGYVLIHNVQVQVPLQRLRYRGTLQFEDYVALVLDNG 120
Db      61 ELTYLPTNASLFLDIOEVQGYVLIHNVQVQVPLQRLRYRGTLQFEDYVALVLDNG 120
Qy      121 DPLNNTPTGASPGALRELORSLETILKGSVLIQRNPOLCYODTILMKDIFHNQOLA 180
Db      121 DPLNNTPTGASPGALRELORSLETILKGSVLIQRNPOLCYODTILMKDIFHNQOLA 180
Qy      181 LTLIDNRSRACHPCSPCKGRCWGESSEDCQSILRTVCAGGACRCKGLPTDCHEQC 240
Db      181 LTLIDNRSRACHPCSPCKGRCWGESSEDCQSILRTVCAGGACRCKGLPTDCHEQC 240
Qy      241 AAGCTGPRGSDCLALHNHSGICELHCPALVTYNTDFFESMPNEGYTGAACVTAAP 300
Db      241 AAGCTGPRGSDCLALHNHSGICELHCPALVTYNTDFFESMPNEGYTGAACVTAAP 300
Qy      301 YNYLSTDVGSCTLVCPLEHNOEVTAEADGTQRCCKSKPCARVCYGLMQYIRANSKFIGIT 360
Db      301 YNYLSTDVGSCTLVCPLEHNOEVTAEADGTQRCCKSKPCARVCYGLMQYIRANSKFIGIT 360
Qy      361 ELFPAGCKKIFGSLAFPESEFDGPASTATLQEOLOVFTLEITGYLIYSAMPDPLP 420
Db      361 IQEFAGCKKIFGSLAFPESEFDGPASTATLQEOLOVFTLEITGYLIYSAMPDPLP 420
Qy      421 DLSVFQNLQVIRGRILHNHGAYSLTLOGIGISWMLGRSLRELGSGLALIHNTHLCPVHTV 480
Db      421 DLSVFQNLQVIRGRILHNHGAYSLTLOGIGISWMLGRSLRELGSGLALIHNTHLCPVHTV 480
Qy      481 PMDQLFNPHQALLHTANRPDECVGSGLAHQICARGHCWGPGPTQCVNCSQFLRQEC 540
Db      481 PMDQLFNPHQALLHTANRPDECVGSGLAHQICARGHCWGPGPTQCVNCSQFLRQEC 540
Qy      541 VEECRVLOGLPREYVNAHCLPCHPECOPOKGSVTCGPREADQCVAAHAKDPFCVARC 600
Db      541 VEECRVLOGLPREYVNAHCLPCHPECOPOKGSVTCGPREADQCVAAHAKDPFCVARC 600

```



```

QY 601 PSQVPLDLYMPIMKPFDEEGACOPCPINCTHSCVDLDDKCPAEQASPLTIVSAVVG 660
DB 601 PSQVPLDLYMPIMKPFDEEGACOPCPINCTHSCVDLDDKCPAEQASPLTIVSAVVG 660
QY 661 ILLVVLGVVFGILIKRQCKIRKYMRLLOETELVEPLTSGAMPNQAKMILKEFNN 720
DB 661 ILLVVLGVVFGILIKRQCKIRKYMRLLOETELVEPLTSGAMPNQAKMILKEFNN 719
QY 721 FTYSFWMRVPKVSAS-HLETYYKGIWPDGENYKIPAIIVLLENTPKANKKILDEAVY 779
DB 720 -----LRKVKVLSSGAFGTVYKGIWPDGENYKIPAIIVLLENTPKANKKILDEAVY 773
QY 780 MAGVSPYVSRLLGICLTSTVOLVTLMPYCLLDHYENRGRGLGSDLLNMCQIAKGN 839
DB 774 MAGVSPYVSRLLGICLTSTVOLVTLMPYCLLDHYENRGRGLGSDLLNMCQIAKGN 833
QY 840 SYLEDVLAHRDLAARVLYKSPHVKITDPGLARLLDDETYHADGKVPYIKMMALES 869
DB 834 SYLEDVLAHRDLAARVLYKSPHVKITDPGLARLLDDETYHADGKVPYIKMMALES 893
QY 900 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVYK 959
DB 894 ILRRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLEKGERLPQPICTIDVYK 953
QY 960 IMVKCMWIDSECRPFRELYSESRMARDPQRFVYIQLNEDLGSPASPLDSTFYRSLLEDD 1019
DB 954 IMVKCMWIDSECRPFRELYSESRMARDPQRFVYIQLNEDLGSPASPLDSTFYRSLLEDD 1013
QY 1020 MGDLVDAEYLVPQGGFCFDPAPAGAGMYHHRSSSTSGGGDLTLGLEPSEEBAPRS 1079
DB 1014 MGDLVDAEYLVPQGGFCFDPAPAGAGMYHHRSSSTSGGGDLTLGLEPSEEBAPRS 1073
QY 1080 PLASEGAGSDVEDDGLGMAAKGLQSLPTHDSPLQRYSEDPVLPSESTDGYVAPLTC 1139
DB 1074 PLASEGAGSDVEDDGLGMAAKGLQSLPTHDSPLQRYSEDPVLPSESTDGYVAPLTC 1133
QY 1140 SPOREYVNOQDVAPQSPREGRPLPARPAGATLERKTLSPKNGVYKQVAFGAVEN 1199
DB 1134 SPOREYVNOQDVAPQSPREGRPLPARPAGATLERKTLSPKNGVYKQVAFGAVEN 1193
QY 1200 PEYLTPQGAAPQHPAPSPAFDNLVYWDODPPEKAPSTFKGPTAENPEYGLDV 1259
DB 1194 PEYLTPQGAAPQHPAPSPAFDNLVYWDODPPEKAPSTFKGPTAENPEYGLDV 1253
QY 1260 PV 1261
DB 1254 PV 1255

RESULT 7
AAB21198
ID AAB21198 standard; protein: 1255 AA.
XX
AC AAB21198;
XX
DT 12-JAN-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Human; HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
XX
OS Homo sapiens.
XX
PN MO200044899-A1.
XX
PD 03-AUG-2000.
XX
PF 28-JAN-2000; 2000MO-US02164.
XX
PR 29-JAN-1999; 99US-0117976.
XX

```

```

PA (CORI-) CORIXA CORP.
PA (SMK) SMITHKLINE BEECHAM.
PI Cheever MA, Cheysen D;
XX WPI; 2000-505976/45.
DR N-PSDB; AAA89736.
XX
PT HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers.
XX
PS Claim 52; Fig 7; 128pp; English.
XX
CC The present sequence is the human HER-2/neu protein. It is a member of
CC the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.
XX
SQ Sequence 1255 AA;
Query Match 96.9%; Score 6641; DB 21; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
QY 1 MELALCRWGLLALLPFGAASVYCTGTDMKLRPASPEHLDMLRLHYQSCQVVGNTL 60
DB 1 MELALCRWGLLALLPFGAASVYCTGTDMKLRPASPEHLDMLRLHYQSCQVVGNTL 60
QY 61 ELTYPTNASSFLDDIOVGQYVLIANNQVQVPLQLRIYRGTLQLEEDNYALAVLNDG 120
DB 61 ELTYPTNASSFLDDIOVGQYVLIANNQVQVPLQLRIYRGTLQLEEDNYALAVLNDG 120
QY 121 DLNNTTPTATGSPGLRELORSLETILKGVLIQRNPOLCYDTIIMKIDFHGNOLA 180
DB 121 DLNNTTPTATGSPGLRELORSLETILKGVLIQRNPOLCYDTIIMKIDFHGNOLA 180
QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSILRTVCAGGACRCKGPLETDCHEQC 240
DB 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSILRTVCAGGACRCKGPLETDCHEQC 240
QY 241 AAGCTGPRXSDCLACHFNHSGICELHCPALVTYVTDTFEEMPNREGYTCASCVTACP 300
DB 241 AAGCTGPRXSDCLACHFNHSGICELHCPALVTYVTDTFEEMPNREGYTCASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPILNNOEYTAEDGTQRCCKSPCARVYCGLMQYIKANSKFIGIT 360
DB 301 YNYLSTDVGSCTLVCPILNNOEYTAEDGTQRCCKSPCARVYCGLMQYIKANSKFIGIT 360
QY 361 ELFEAGCKKIFGSLAFLESPFGDPASTAPLQEPQLOVFFLEITIGVLIISAMPDSL 420
DB 361 ELFEAGCKKIFGSLAFLESPFGDPASTAPLQEPQLOVFFLEITIGVLIISAMPDSL 420
QY 421 DLSTVFQNIQVIRGLIHNGAVSLTQIGISIMLGIRSLRELSGSLALHHNTHLCFVTV 480
DB 421 DLSTVFQNIQVIRGLIHNGAVSLTQIGISIMLGIRSLRELSGSLALHHNTHLCFVTV 480
QY 481 FWDQLFRNPHQALLHTANRPEDECVGEGIALCHQLCARHGWGPGFTQCVCNSQFLRGEC 540
DB 481 FWDQLFRNPHQALLHTANRPEDECVGEGIALCHQLCARHGWGPGFTQCVCNSQFLRGEC 540
QY 541 VEEGVVLQGLPREYVNAHCHLPCHPECCPQNGSVTCFPEEDQCACAHYKDPPECVARC 600
DB 541 VEEGVVLQGLPREYVNAHCHLPCHPECCPQNGSVTCFPEEDQCACAHYKDPPECVARC 600
QY 601 PSQVPLDLYMPIMKPFDEEGACOPCPINCTHSCVDLDDKCPAEQASPLTIVSAVVG 660

```

```

Db      601 PSQVAPLSTWPIKPFDEBGAQCPINCTHSCVDLDDGCEPCQASPLTSTISAVVG 660
QY      661 ILLVVLGVVFGILLIKRQOKIRKYTRRLLOETELVEPLTPSGAMNQAKMILKEFNN 720
Db      661 ILLVVLGVVFGILLIKRQOKIRKYTRRLLOETELVEPLTPSGAMNQAKMILKEFNN 719
QY      721 FTYSFMLRVKVSNS-HLETVYKGIWIPDGENVKIPALIVALENTSPKANKILDEAYV 779
Db      720 -----LRKVKVSGAGFTVYKGIWIPDGENVKIPALIVALENTSPKANKILDEAYV 773
QY      780 MAGVSPYVSRLLGICLTSTVQVLTQMLPYGCLLDHRENRGRGLSGODLLMCMQIAKGM 839
Db      774 MAGVSPYVSRLLGICLTSTVQVLTQMLPYGCLLDHRENRGRGLSGODLLMCMQIAKGM 833
QY      840 SYLEDVTLVHRDLAARVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMMALES 899
Db      834 SYLEDVTLVHRDLAARVLYKSPNHVKITDFGLARLLDIDETEVHADGKVPKIMMALES 893
QY      900 ILRRFTHSDVNSYGVTVWELMTFGAKPYDGI PAHEIPDLLEKGRLPQPICTIDVYM 959
Db      894 ILRRFTHSDVNSYGVTVWELMTFGAKPYDGI PAHEIPDLLEKGRLPQPICTIDVYM 953
QY      960 IMYKCMWIDSECRPRFRELVSFESRMAKDQRFVVIQNEDLGPASPLDSTFYRLLEDD 1019
Db      954 IMYKCMWIDSECRPRFRELVSFESRMAKDQRFVVIQNEDLGPASPLDSTFYRLLEDD 1013
QY      1020 MGLVDAEETLVQCGFCPPDPAPAGAGMTHHRSSSTSSGGGDLTLGLEPSEBAPRS 1079
Db      1014 MGLVDAEETLVQCGFCPPDPAPAGAGMTHHRSSSTSSGGGDLTLGLEPSEBAPRS 1073
QY      1080 PLAPSEGAGSDVFDGLGMAAKGLQSLPTHDSPLQRYSEDTVPLPSETDXYVAPLTC 1139
Db      1074 PLAPSEGAGSDVFDGLGMAAKGLQSLPTHDSPLQRYSEDTVPLPSETDXYVAPLTC 1133
QY      1140 SPPEYVNOVDVAPQPPSPREGPLPAARPAAGATLEPRAKTLSPKNGVYKQVAFGAVEN 1199
Db      1134 SPPEYVNOVDVAPQPPSPREGPLPAARPAAGATLEPRAKTLSPKNGVYKQVAFGAVEN 1193
QY      1200 PEYLTPQGAAPQPPSPAFDMLYYWDODPPERGAPESTFKCTPTAENPEVYGLDV 1259
Db      1194 PEYLTPQGAAPQPPSPAFDMLYYWDODPPERGAPESTFKCTPTAENPEVYGLDV 1253
QY      1260 PV 1261
Db      1254 PV 1255

RESULT 8
AA84780
ID      AAY84780 standard; Protein; 1255 AA.
XX
AC      AAY84780;
XX
DT      08-AUG-2000 (first entry)
XX
DE      Amino acid sequence of the SPLICE erbb-2 receptor protein.
XX
KW      SPLICE erbb-2 receptor protein; cell transformation disorder; cancer;
KW      tumor cell proliferation; tissue degeneration; arthropathy;
KW      bone resorption; inflammatory disease; degenerative disorder;
XX
OS      Homo sapiens.
XX
PN      WO200020579-A1.
XX
PD      13-APR-2000.
XX
PF      01-OCT-1999; 99WO-CA00912.
XX
PR      02-OCT-1998; 98US-0165192.
XX

```

```

PA      (UNMC-) UNIV MCMMASTER.
XX
PI      Muller WJ, Siegel PM,
XX
DR      WPI: 2000-303768/26.
XX
DR      N-PSDB; AAA14812.
XX
PT      Nucleic acid encoding an erbb 2 receptor protein designated SPLICE
XX
PT      erbb-2, inhibitors of the protein are useful for treatment of cancer -
XX
PS      Claim 3; Fig 2; 60pp; English.
XX
CC      The present sequence represents a SPLICE erbb-2 receptor protein. The
CC      protein has an in-frame deletion of 16 amino acids, 2 of which are
CC      conserved cysteine residues, compared to the unspliced protein. The
CC      erbb-2 polynucleotide is used to construct probes for detecting
CC      disorders of cell transformation such as cancer. Antibodies to the
CC      protein may be used to detect SPLICE erbb-2 in a sample. Agents
CC      (e.g. antisense oligonucleotides) which inhibit the expression of
CC      SPLICE erbb-2 are useful for reducing tumor cell proliferation and
CC      treating cancer. Substances which stimulate SPLICE erbb-2 are useful
CC      for treating conditions involving damaged cells including conditions
CC      in which degeneration of tissue occurs, such as arthropathy, bone
CC      resorption, inflammatory diseases, degenerative disorders of the
CC      central nervous system and wound healing.
XX
SQ      Sequence 1255 AA;
XX
Query Match 96.9%; Score 6641; DB 21; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
QY      1 MEALALCWGILLALPPGAASVQCTGTMKLRIPASPEHLDMLRLHYOGCQVQGNL 60
Db      1 MEALALCWGILLALPPGAASVQCTGTMKLRIPASPEHLDMLRLHYOGCQVQGNL 60
QY      61 ELTYPTVASLSFLDIDIEVQGYVLIANOVROYPLQRLTVRGQLFEDNYALAVLDNG 120
Db      61 ELTYPTVASLSFLDIDIEVQGYVLIANOVROYPLQRLTVRGQLFEDNYALAVLDNG 120
QY      121 DPLNNTPTVGTASPGGLRELQRLSTLEILKGVLIQRNPQLCYODTILMKDIFHNQOLA 180
Db      121 DPLNNTPTVGTASPGGLRELQRLSTLEILKGVLIQRNPQLCYODTILMKDIFHNQOLA 180
QY      181 LTLIDTNSRACHPESPMCKSRCKGSESECCQSLTRVCGAGCARCKGRLPTDCHEQC 240
Db      181 LTLIDTNSRACHPESPMCKSRCKGSESECCQSLTRVCGAGCARCKGRLPTDCHEQC 240
QY      241 AAGCTGPRGSDCLAFHNSGICELHPALVTVNTDPEFMPNDEGRYTGCASCVTACP 300
Db      241 AAGCTGPRGSDCLAFHNSGICELHPALVTVNTDPEFMPNDEGRYTGCASCVTACP 300
QY      301 YNYLSTDVGSCTLVCPHNOEVTADGTORCEKSKPCARVYCYGLGMQYIKANSKFIGIT 360
Db      301 YNYLSTDVGSCTLVCPHNOEVTADGTORCEKSKPCARVYCYGLGMQYIKANSKFIGIT 360
QY      361 ELEFAGCKKIRGSLAFIPESFDPGPA NSTAIOEOLQVETLEITGYLIASMPDPLP 420
Db      361 IQEPAGCKKIRGSLAFIPESFDPGPA NSTAIOEOLQVETLEITGYLIASMPDPLP 420
QY      421 DLSVFQMLQVIRGRILHNAYSLTLQIGISWLGRLRLRELSGLALIHNTHTLCFYHTV 480
Db      421 DLSVFQMLQVIRGRILHNAYSLTLQIGISWLGRLRLRELSGLALIHNTHTLCFYHTV 480
QY      481 PMDOLFRNPHALHTANRPEDEVGGLACHQICARGHCKGPGPTCCVNCOSQLRQEC 540
Db      481 PMDOLFRNPHALHTANRPEDEVGGLACHQICARGHCKGPGPTCCVNCOSQLRQEC 540
QY      541 VEECRVQLPREYVNAARHCLPCHPECOPOGNGSYTCGPEADQCVACAHYKDPFFCYARC 600
Db      541 VEECRVQLPREYVNAARHCLPCHPECOPOGNGSYTCGPEADQCVACAHYKDPFFCYARC 600
QY      601 PSQVAPLSTWPIKPFDEBGAQCPINCTHSCVDLDDGCEPCQASPLTSTISAVVG 660

```



```

Db 601 PSQVPSDLSYMPIMKFPDEEGACQPCPINCSTHSCVDLDDKCGPAEQRASPLTISIISAVG 660
Qy 661 ILLVVVGVVFGILLIKRQOKIRKXTMRLLLOETELVEPLTPSGAMPQAOQMRLLKEENN 720
Db 661 ILLVVVGVVFGILLIKRQOKIRKXTMRLLLOETELVEPLTPSGAMPQAOQMRLLKEETE- 719
Qy 721 FTVSFMLEVPKVSAS-NLETVYKGIWIPDGENVKIPVAIKVLENTSEKANKELIDEAIV 779
Db 720 -----LKKVKVLGSGAFGTYYKGIWIPDGENVKIPVAIKVLENTSEKANKELIDEAIV 773
Qy 780 MAGVSPVYSRLILGICLTSTVQVLTQMLPFGCLLDHRENGRLGSDLLNMCQIAKGM 839
Db 774 MAGVSPVYSRLILGICLTSTVQVLTQMLPFGCLLDHRENGRLGSDLLNMCQIAKGM 833
Qy 840 SYLEDVRLVHRDLAARNLVKSPNHKITTDFGLARLIDIDEETHADGKVPIMKMALES 899
Db 834 SYLEDVRLVHRDLAARNLVKSPNHKITTDFGLARLIDIDEETHADGKVPIMKMALES 893
Qy 900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 959
Db 894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKGERLPQPPICITIDVYM 953
Qy 960 IMVKCMTIDSECRPFRELVSEFSMARDPQRFVYIQNEDLGPAASPLDSTFYRLLEDDE 1013
Db 954 IMVKCMTIDSECRPFRELVSEFSMARDPQRFVYIQNEDLGPAASPLDSTFYRLLEDDE 1013
Qy 1020 MGDLVDAEYLVPOQGFPCPDPAFAGAGMVRHRRSSSTRSGGDLTLGLEPSEBEAPRS 1079
Db 1014 MGDLVDAEYLVPOQGFPCPDPAFAGAGMVRHRRSSSTRSGGDLTLGLEPSEBEAPRS 1073
Qy 1080 PLABEGAGSDVFDGDLGMAKAGLSLPTHDPSPLOKYSDDPYVPLSEIDGYVAPLTC 1139
Db 1074 PLABEGAGSDVFDGDLGMAKAGLSLPTHDPSPLOKYSDDPYVPLSEIDGYVAPLTC 1133
Qy 1140 SPOEYVNPQDVRPQPSPREGRLPAARPAAGATLERAATLSFGKNGVVKOVFAFGAVEN 1199
Db 1134 SPOEYVNPQDVRPQPSPREGRLPAARPAAGATLERAATLSFGKNGVVKOVFAFGAVEN 1193
Qy 1200 PEYLTPQGAAPQHPHPPAFSPADNLYYWDODPPERGAPSTFKGTFTANPEYGLDV 1259
Db 1194 PEYLTPQGAAPQHPHPPAFSPADNLYYWDODPPERGAPSTFKGTFTANPEYGLDV 1253
Qy 1260 PV 1261
Db 1254 PV 1255

RESULT 9
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
PN MO200153463-A2.
XX
PD 26-JUL-2001.
XX
PE 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
PA (CORI-) CORIXA CORP.
XX

```

```

PI Cheever MA, Hand-Zimmermann S;
XX
DR WPI: 2001-476112/51.
XX N-PSDB; AAB23392.
XX
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer.
XX
PS Claim 2; Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;
XX
Query Match 96.3%; Score 6641; DB 22; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
Qy 1 METALCRMGILLALPPGAASVQVCTGDMKRLPASPETHLDMRLHYOGQVQGNL 60
Db 1 METALCRMGILLALPPGAASVQVCTGDMKRLPASPETHLDMRLHYOGQVQGNL 60
Qy 61 ELTYLPTNASSLSLQIDIEVQGYVLIANNQVRVPLQRLIRVGTQLFEDNVALAVLNG 120
Db 61 ELTYLPTNASSLSLQIDIEVQGYVLIANNQVRVPLQRLIRVGTQLFEDNVALAVLNG 120
Qy 121 DPLNNTPTVAGAPGLRELOSLTEILKGGVLIQRNQLCYQDITLWKDIFHKNNOLA 180
Db 121 DPLNNTPTVAGAPGLRELOSLTEILKGGVLIQRNQLCYQDITLWKDIFHKNNOLA 180
Qy 181 LTLIDNRSBACHPCSPMKSGRCWGESSEDCQSLTRTVACAGCARKGPLEPTCCHEQC 240
Db 181 LTLIDNRSBACHPCSPMKSGRCWGESSEDCQSLTRTVACAGCARKGPLEPTCCHEQC 240
Qy 241 AAGCTGPKSDCLACHFNHSGICEHLCALVTYNTDTRESNPNBGRRTFFGASCTYAC 300
Db 241 AAGCTGPKSDCLACHFNHSGICEHLCALVTYNTDTRESNPNBGRRTFFGASCTYAC 300
Qy 301 YNYLSTDVSGCTVCPLEHNOEYTAEDGTORCEKSPKARVCYGLCMOYIKANSKFIGIT 360
Db 301 YNYLSTDVSGCTVCPLEHNOEYTAEDGTORCEKSPKARVCYGLCMOYIKANSKFIGIT 360
Qy 361 ELEFPAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFTLEBITGLYISAWPDSL 420
Db 361 ELEFPAGCKKIFGSLAFLPESFDGDPASNTAPLOPELOVFTLEBITGLYISAWPDSL 420
Qy 421 DLSVFQNLQVIRGRIHNGAYSLTLOGLSIMLGLRSLRELSGLALIHNNHLCGVHIV 480
Db 421 DLSVFQNLQVIRGRIHNGAYSLTLOGLSIMLGLRSLRELSGLALIHNNHLCGVHIV 480
Qy 481 PMDQFRNPHQALLHTANRPEDECVGEGJACHQLCARGHCWGGPTQVNCQGFJRGQEC 540
Db 481 PMDQFRNPHQALLHTANRPEDECVGEGJACHQLCARGHCWGGPTQVNCQGFJRGQEC 540
Qy 541 VEECRVLOGLPREYVNAHRCLECHPECCQONSVCYTFGBAQQVACAHYKXPPFVAC 600
Db 541 VEECRVLOGLPREYVNAHRCLECHPECCQONSVCYTFGBAQQVACAHYKXPPFVAC 600
Qy 601 PSQVPSDLSYMPIMKFPDEEGACQPCPINCSTHSCVDLDDKCGPAEQRASPLTISIISAVG 660
Db 601 PSQVPSDLSYMPIMKFPDEEGACQPCPINCSTHSCVDLDDKCGPAEQRASPLTISIISAVG 660
Qy 661 ILLVVVGVVFGILLIKRQOKIRKXTMRLLLOETELVEPLTPSGAMPQAOQMRLLKEENN 720

```

Db 661 ILVVVLGVVFGILIKRRQKIKYTKRLLOETELVEPLTPSGAMPNQAQNRILKETE 719  
 Qy 721 FTVSFWLVRPKVSAS-HLETVYGIWIPDGENYKIPVAIKVRENTSPKANKEILDEAVY 779  
 Db 720 -----LRKVKVLGSGAFGTYYKGIWIPDGENYKIPVAIKVRENTSPKANKEILDEAVY 773  
 Qy 780 MAGVSPYVRLIGLICITSTVOLTQVLTQMPYGCILDVHRENRKGLSGSDLLNMOMQIAKGM 839  
 Db 774 MAGVSPYVRLIGLICITSTVOLTQVLTQMPYGCILDVHRENRKGLSGSDLLNMOMQIAKGM 833  
 Qy 840 SYLEDVRLVHRDLAARVNLVKSNNHVKITDFGLARLLDIDETEHYHAGDGKVPKIMMALES 899  
 Db 834 SYLEDVRLVHRDLAARVNLVKSNNHVKITDFGLARLLDIDETEHYHAGDGKVPKIMMALES 893  
 Qy 900 ILRRRFTHSDVMSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGRLLPQPICTIDVYM 959  
 Db 894 ILRRRFTHSDVMSYGVTTWELMTFGAKPYDGIIPAEIPDLLEKGRLLPQPICTIDVYM 953  
 Qy 960 IMKCMWIDSECRPRFRELVSFESRMARDPQRFVVIQNEPLGPAFLDSFYSRLLEDD 1019  
 Db 954 IMKCMWIDSECRPRFRELVSFESRMARDPQRFVVIQNEPLGPAFLDSFYSRLLEDD 1013  
 Qy 1020 MGDLVDAEYLVVQGGFCPPDPAFGAGVNHRRSSSTRSGGDLTLGLBSEBAPRS 1079  
 Db 1014 MGDLVDAEYLVVQGGFCPPDPAFGAGVNHRRSSSTRSGGDLTLGLBSEBAPRS 1073  
 Qy 1080 PLAPSGAGSDVFDGDLGGAAGKLSLPHDPSLQKRSSEPTVPLBETDGYAPLTC 1139  
 Db 1074 PLAPSGAGSDVFDGDLGGAAGKLSLPHDPSLQKRSSEPTVPLBETDGYAPLTC 1133  
 Qy 1140 SPQEVYVNPQVAPQPPSPREGLPAPAPAGATLERAKTLSPGKNGVNDVAFAGAVEN 1199  
 Db 1134 SPQEVYVNPQVAPQPPSPREGLPAPAPAGATLERAKTLSPGKNGVNDVAFAGAVEN 1193  
 Qy 1200 PEYLTPQGAAPQPPHPPAPSPAFDNLVYWDQPPERGAAPSTFKTPTAENPEYIGLDV 1259  
 Db 1194 PEYLTPQGAAPQPPHPPAPSPAFDNLVYWDQPPERGAAPSTFKTPTAENPEYIGLDV 1253  
 Qy 1260 PV 1261  
 Db 1254 PV 1255

RESULT 10  
 AA088267  
 ID AA088267 standard; Protein; 1255 AA.  
 AC AA088267;  
 XX  
 DT 11-SEP-2001 (first entry)  
 XX  
 DE HER2/neu amino acid sequence.  
 XX  
 KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;  
 KW immune response; vaccine; tumour; cancer; cytotoxic; immunostimulant;  
 KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.  
 XX  
 OS Homo sapiens.  
 PN WO200141787-A1.  
 PD 14-JUN-2001.  
 PF 11-DEC-2000; 2000MO-US33591.  
 PR 10-DEC-1999; 99US-0458299.  
 PA (EPIM-) EPIMUNE INC.  
 PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;  
 DR WPI, 2001-374995/39.

XX An isolated prepared HER2/neu epitope useful in a vaccine for inducing  
 PT cellular immune responses for the prevention and treatment of cancer  
 XX  
 PS Disclosure, Page 15; 199pp; English.

CC The present invention describes isolated prepared HER2/neu epitopes (I).  
 CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is  
 CC culture in vitro and binds to a complex of an epitope (II), bound to a  
 CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)  
 CC and a second epitope and the peptide is less than 50 contiguous amino  
 CC acids that have 100% identity with a native peptide sequence of HER2/neu;  
 CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical  
 CC excipient; (4) an isolated nucleic acid encoding a peptide comprising  
 CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytotoxic  
 CC and immunostimulant activities, and can be used in vaccines. (I), (II)  
 CC and (III) are useful for inducing cellular immune responses for the  
 CC prevention and treatment of cancer. (I) and (II) are useful for  
 CC monitoring or evaluating an immune response to a tumour-associated  
 CC antigen when incubated with a T lymphocyte sample from a patient and  
 CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope  
 CC based vaccines mean that immunosuppressive epitopes that may be present  
 CC in whole antigens may be avoided. Selected epitopes may be combined to  
 CC enhance immunogenicity. The possible pathological side effects caused by  
 CC infectious agents or whole protein antigen is eliminated. The vaccine  
 CC provides the ability to direct and focus an immune response to multiple  
 CC selected antigens from the same pathogen. Epitope-based anti-tumour  
 CC vaccines provides the opportunity to combine epitopes derived from  
 CC multiple tumour-associated molecules addressing the problem of tumour-  
 CC tumour variability and reducing the likelihood of tumour escape due to  
 CC antigen loss. AA088266 to AA089121 represent amino acid sequences used in  
 CC the exemplification of the present invention.

Sequence 1255 AA;

Query Match 96.9%; Score 6641; DB 22; Length 1255;  
 Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;

Qy 1 MELALCRWGLLLALPPGAASVCTGTDMKRLPASPETHLMDLRLHYGCGVOVQNTL 60  
 Db 1 MELALCRWGLLLALPPGAASVCTGTDMKRLPASPETHLMDLRLHYGCGVOVQNTL 60  
 Qy 61 ELTYLPTNASLFIQDIQEVGVYLIANQVQVPLQRLIRVGTQLEFNYALAVLDNG 120  
 Db 61 ELTYLPTNASLFIQDIQEVGVYLIANQVQVPLQRLIRVGTQLEFNYALAVLDNG 120  
 Qy 121 DPLNNTTPVTGASPGGRELQRLSTELKGVLIQBNPOLCYGDTLLMKDIFFKNNOLA 180  
 Db 121 DPLNNTTPVTGASPGGRELQRLSTELKGVLIQBNPOLCYGDTLLMKDIFFKNNOLA 180  
 Qy 121 DPLNNTTPVTGASPGGRELQRLSTELKGVLIQBNPOLCYGDTLLMKDIFFKNNOLA 180  
 Db 121 DPLNNTTPVTGASPGGRELQRLSTELKGVLIQBNPOLCYGDTLLMKDIFFKNNOLA 180  
 Qy 181 LTLIDYNSRACHPCSPCKGSRGSGESSDDQSLRTVACAGGACRKGPLPTDCHEQC 240  
 Db 181 LTLIDYNSRACHPCSPCKGSRGSGESSDDQSLRTVACAGGACRKGPLPTDCHEQC 240  
 Qy 241 AAGCTGPKHSDCLALCPHNSGICELCPALVYNTDTEFSMPRPERYTFGASCVAACP 300  
 Db 241 AAGCTGPKHSDCLALCPHNSGICELCPALVYNTDTEFSMPRPERYTFGASCVAACP 300  
 Qy 301 YNYLSTDVGSCTIVCPAHNOEVAEDGTQCEKSKPCARVCYGLQVYIKANSKFIGIT 360  
 Db 301 YNYLSTDVGSCTIVCPAHNOEVAEDGTQCEKSKPCARVCYGLQVYIKANSKFIGIT 360  
 Qy 361 ELERAGCKTFGSLAFIPESFDGDPASNTAPLOBOLQVETLEITGYLYISAMPDLP 420  
 Db 361 ELERAGCKTFGSLAFIPESFDGDPASNTAPLOBOLQVETLEITGYLYISAMPDLP 420  
 Qy 421 DLSYFQNLQVIRGRILNAGVSLTLQGLGSMGLSRLBGLGLIHNHTLCFHTY 480  
 Db 421 DLSYFQNLQVIRGRILNAGVSLTLQGLGSMGLSRLBGLGLIHNHTLCFHTY 480  
 Qy 481 PMDQLFNPQIALHNPEDQVGBGLACHQLCARGHCWGPPTQCVNCSOPLRGQEC 540  
 Db 481 PMDQLFNPQIALHNPEDQVGBGLACHQLCARGHCWGPPTQCVNCSOPLRGQEC 540

Db 481 FMDOLFRRPHQALHTANRPEDCEVGEGLACHOLCARHCHWGPGPTQCVNCSQFLRGEC 540  
 Qy 541 VEEGRVVOGLPREYVNAHCLPCHPECOFONGSVTCGPPADOCVACAHXKDPFCVARC 600  
 Db 541 VEEGRVVOGLPREYVNAHCLPCHPECOFONGSVTCGPPADOCVACAHXKDPFCVARC 600  
 Qy 601 PSYVCPDLSTYMPWKPFDEEGACPCPINCTHSCVDLDDKCPAEQASPLTJISAVVG 660  
 Db 601 PSYVCPDLSTYMPWKPFDEEGACPCPINCTHSCVDLDDKCPAEQASPLTJISAVVG 660  
 Qy 661 ILVAVVVGAVFGILIKRROOKIRKRYTRRLLOETELVEPLTSGAMNOMKILKEFN 720  
 Db 661 ILVAVVVGAVFGILIKRROOKIRKRYTRRLLOETELVEPLTSGAMNOMKILKEFN 719  
 Qy 721 FTVSFWLRVPRVSAH-HLETYKGIWIPDENVKIPVAIKVLEBNTSPKANKELDEAYV 779  
 Db 720 -----LRKVVLGSAFGTYKGIWIPDENVKIPVAIKVLEBNTSPKANKELDEAYV 773  
 Qy 780 MAGVSPYVSRLLGICITSTVQVLTQMPYGCILDHVRNRRGLSGDILLNMCOTAKGM 839  
 Db 774 MAGVSPYVSRLLGICITSTVQVLTQMPYGCILDHVRNRRGLSGDILLNMCOTAKGM 833  
 Qy 840 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLIDETEXHADGKVPKIMMALES 899  
 Db 834 SYLEDVRLVHRDLAARNVLYKSPNHVKITDFGLARLIDETEXHADGKVPKIMMALES 893  
 Qy 900 ILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGI-PAREIPDLEKEBERLPQPICTIDVYM 959  
 Db 894 ILRRRFTHOSDVMSYGVTVWELMTFGAKPYDGI-PAREIPDLEKEBERLPQPICTIDVYM 953  
 Qy 960 IMVKCMMIDSECRFRPELSESEFRMARDPORFVIONEDLGPASPLDSTFYRSLLEDDO 1019  
 Db 954 IMVKCMMIDSECRFRPELSESEFRMARDPORFVIONEDLGPASPLDSTFYRSLLEDDO 1013  
 Qy 1020 MGDVDAEELVLPQGFCEPDPAFGAGMWHRRSSSTRSGGDLTLGLPESSEBAPRS 1079  
 Db 1014 MGDVDAEELVLPQGFCEPDPAFGAGMWHRRSSSTRSGGDLTLGLPESSEBAPRS 1073  
 Qy 1080 PLAPSEGASDVPEPGDILGMAAKGLOSLPTHDPSPLORYSEDEPTVLPSTDDVAPLTC 1139  
 Db 1074 PLAPSEGASDVPEPGDILGMAAKGLOSLPTHDPSPLORYSEDEPTVLPSTDDVAPLTC 1133  
 Qy 1140 SPOEYVNOVDVSPQSPREBGLPAARPAAGATLERAKTILSPKNGVVKDVFAGAVEN 1199  
 Db 1134 SPOEYVNOVDVSPQSPREBGLPAARPAAGATLERAKTILSPKNGVVKDVFAGAVEN 1193  
 Qy 1200 PEYLTPOGGAAPQHPPPASPAFDNLYWDQPPERGAPOSTFKGTPTANPEYVGLDV 1259  
 Db 1194 PEYLTPOGGAAPQHPPPASPAFDNLYWDQPPERGAPOSTFKGTPTANPEYVGLDV 1253  
 Qy 1260 PV 1261  
 Db 1254 PV 1255  
 RESULT 11  
 AAE24067  
 ID AAE24067 standard; Protein: 1255 AA.  
 XX AAE24067;  
 DT 23-SEP-2002 (first entry)  
 XX Human Her-2 protein.  
 XX Human Her-2 protein.  
 XX Human; Her-2; epidermal growth factor receptor 2; infection; cancer;  
 KM hyperproliferative disorder; prophylaxis; inflammation; antisense;  
 KM tumour; gene therapy; phosphorocinicate backbone.  
 XX Homo sapiens.  
 OS  
 PN MO200222636-A1.  
 XX

PD 21-MAR-2002.  
 XX 12-SEP-2001; 2001MO-US28572.  
 PF 15-SEP-2000; 2000US-0663834.  
 PR (ISIS-) ISIS PHARM INC.  
 PA Bennett CF, Cowse LM;  
 PI MPI: 2002-471192/50.  
 DR N-PSDB; AAD38904.  
 XX  
 PT Novel antisense oligonucleotide which modulates the expression of Human  
 PT Epidermal Growth Factor receptor, Her2, is useful for treating tumors  
 PT inflammation or to prevent infection in humans -  
 XX  
 PS Example 13; Page 95-107; 116pp; English.  
 CC The invention relates to antisense compounds targeted to a nucleic  
 CC acid molecule encoding Her2 (human Epidermal Growth Factor receptor 2)  
 CC that specifically hybridizes with and inhibits the expression of Her2.  
 CC Antisense compounds of the invention are used for treating diseases or  
 CC conditions associated with Her2 such as hyperproliferative disorders  
 CC e.g. lung, breast, gastric, oesophageal, colon, bladder, salivary,  
 CC neural or cardiac cancer. They are also useful prophylactically e.g.  
 CC to prevent or delay infection, inflammation and tumour formation. The  
 CC invention is also used in gene therapy. The present sequence is human  
 CC Her-2 protein.  
 XX  
 SQ Sequence 1255 AA;  
 Query Match 96.9%; Score 6641; DB 23; Length 1255;  
 Best Local Similarity 97.3%; Pred. No. 0;  
 Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;  
 Qy 1 MELALCWMGLLALLPGASTQVCTGTMKRLPASPETHLWLRRLVYQCCVQVQNL 60  
 Db 1 MELALCWMGLLALLPGASTQVCTGTMKRLPASPETHLWLRRLVYQCCVQVQNL 60  
 Qy 61 ELTYLPTNASSFLDIOEVQGVYLIANOVQVPLQSLRIVRGTLQLEEDVYALAVLNG 120  
 Db 61 ELTYLPTNASSFLDIOEVQGVYLIANOVQVPLQSLRIVRGTLQLEEDVYALAVLNG 120  
 Qy 121 DPLNNTTPTVGTASPGGLRELQRLSLTELKGVLIQNPOLCYQDTILMKDIFHKNOILA 180  
 Db 121 DPLNNTTPTVGTASPGGLRELQRLSLTELKGVLIQNPOLCYQDTILMKDIFHKNOILA 180  
 Qy 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSITRTVCGAGGRCGRLPTDCCHQC 240  
 Db 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSITRTVCGAGGRCGRLPTDCCHQC 240  
 Qy 241 AAGCTGPGHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTAC 300  
 Db 241 AAGCTGPGHSDCLACLFHNSGICELHCPALVTYNTDTFESMPNEGRYTFGASCVTAC 300  
 Qy 301 YNYLSTDVSGCTIVCEPLNCEVTAEDGRCCKSPCARCYGUGVYIANSKFIGIT 360  
 Db 301 YNYLSTDVSGCTIVCEPLNCEVTAEDGRCCKSPCARCYGUGVYIANSKFIGIT 360  
 Qy 361 ELLEFAGCKKIFGSLAFPESEFDGPASNTAFLQEOLOVFETLEITGYLISAMPDLP 420  
 Db 361 IOEFAGCKKIFGSLAFPESEFDGPASNTAFLQEOLOVFETLEITGYLISAMPDLP 420  
 Qy 421 DLSVFQNLQVIRGRILNHGAYSLTQIGISWLGRLSRRELGSGLALHNHTHLCFVHTV 480  
 Db 421 DLSVFQNLQVIRGRILNHGAYSLTQIGISWLGRLSRRELGSGLALHNHTHLCFVHTV 480  
 Qy 481 FMDOLFRRPHQALHTANRPEDCEVGEGLACHOLCARHCHWGPGPTQCVNCSQFLRGEC 540  
 Db 481 FMDOLFRRPHQALHTANRPEDCEVGEGLACHOLCARHCHWGPGPTQCVNCSQFLRGEC 540  
 Qy 541 VEEGRVVOGLPREYVNAHCLPCHPECOFONGSVTCGPPADOCVACAHXKDPFCVARC 600

```

Db      541 VEECRVLDGLPRRYVNAHRLPCHPEQPNQSVTFGPADUCVCAHAKBPFCVAC 600
Qy      601 PGGVDPDLSTYMPIMKFPDEBACQPCPINTHSCVDLDKDGCAERASPLTISVAVG 660
Db      601 PGGVDPDLSTYMPIMKFPDEBACQPCPINTHSCVDLDKDGCAERASPLTISVAVG 660
Qy      661 ILVVVLGVFGILIKRQCKIRKYMRLLOETELVEPLTPSGAMPNOQMILKEFN 720
Db      661 ILVVVLGVFGILIKRQCKIRKYMRLLOETELVEPLTPSGAMPNOQMILKEFN 719
Qy      721 FTVSFMLRVPKVAS-HLETYKGIWIPDGENKIPVAKVRENSPANKELIDEAV 779
Db      720 -----LRKVKVLGSGAFGTIVYKIMIPDGENKIPVAKVRENSPANKELIDEAV 773
Qy      780 MAGVSPYVSRLLGICLTSTVQVLTOLMPYGLLDHVRNKGRLSGODLLNMCQIAKGM 839
Db      774 MAGVSPYVSRLLGICLTSTVQVLTOLMPYGLLDHVRNKGRLSGODLLNMCQIAKGM 833
Qy      840 SYLEDVRLVHRDLAARVLVKSNNHVKITDFGLARLLDIDETEHADGKVPKMALES 899
Db      834 SYLEDVRLVHRDLAARVLVKSNNHVKITDFGLARLLDIDETEHADGKVPKMALES 893
Qy      900 ILRRFTHQSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVYM 959
Db      894 ILRRFTHQSDVMSYGVTWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPICTIDVYM 953
Qy      960 IMKCMWIDECRRPRELSESRMARDPQRFVYQNEBGLGASPLDSTFYSSLLEDD 1019
Db      954 IMKCMWIDECRRPRELSESRMARDPQRFVYQNEBGLGASPLDSTFYSSLLEDD 1013
Qy      1020 MGLDVAEEYLVQOQGFCDPAPAGAGMWHHRSSSTRSGGDLTLGLEPSEERAPRS 1079
Db      1014 MGLDVAEEYLVQOQGFCDPAPAGAGMWHHRSSSTRSGGDLTLGLEPSEERAPRS 1073
Qy      1080 PLAPSGAGSDVFDGDLGMAKAGLSLPTHPSPQRYSEDTVPPLSETDGYVAPLTC 1139
Db      1074 PLAPSGAGSDVFDGDLGMAKAGLSLPTHPSPQRYSEDTVPPLSETDGYVAPLTC 1133
Qy      1140 SPQEVYNQDVNPPSPREBPLPARPAGATLEBPKTLSPKKNVWVDVAFGAVEN 1199
Db      1134 SPQEVYNQDVNPPSPREBPLPARPAGATLEBPKTLSPKKNVWVDVAFGAVEN 1193
Qy      1200 PEYLTPQGAAPQPHPPAPSPAFDNLVYWDQDPERGAQPPSTFKCTPTAENPEYLGLDV 1259
Db      1194 PEYLTPQGAAPQPHPPAPSPAFDNLVYWDQDPERGAQPPSTFKCTPTAENPEYLGLDV 1253
Qy      1260 PV 1261
Db      1254 PV 1255

RESULT 12
AAE20479
ID      AAE20479 standard; Protein; 1255 AA.
XX      AAE20479;
AC      AAE20479;
XX      01-JUL-2002 (first entry)
DT      01-JUL-2002 (first entry)
XX      Human Her-2/neu protein.
DE      Human Her-2/neu protein.
XX      Human; Her-2/Neu protein; immune response; gene therapy; breast cancer;
KW      human leukocyte antigen; HLA; vaccine; malignancy; cytostatic.
XX      Homo sapiens.
OS      Homo sapiens.
XX      Homo sapiens.
FH      Key
FT      Region
XX      Location/Qualifiers
XX      1021..1030
XX      /note="Naturally processed HLA-B44-restricted epitope"
XX      WO200214503-A2.

```

```

PD      21-FEB-2002.
XX      14-AUG-2001; 2001WO-US41733.
PF      14-AUG-2001; 2000US-225152P.
XX      14-AUG-2000; 2000US-225152P.
XX      28-SEP-2000; 2000US-236428P.
PR      21-FEB-2001; 2001US-270520P.
XX      (CORI-) CORIXA CORP.
XX      Hand-zimmermann S, Cheever MA, Foy TM, Lodes MJ, Kalos MD;
PI      McNeill PD, Vedral TS;
XX      WPI: 2002-280758/32.
DR      N-PSDB; AAD32743.
XX      Novel isolated Her-2/Neu polypeptide composition useful for therapy,
PT      prevention and diagnosis of cancer, preferably breast cancer
XX      Disclosure; Page 114-117; 129pp; English.
XX      The invention relates to an isolated Her-2/Neu polypeptide composition
XX      effective for eliciting an immune response. The invention is useful for
XX      eliciting an immune response in a patient, where the patient is human
XX      leukocyte antigen (HLA)-B44 positive or is affected with breast cancer.
XX      The composition is useful for the therapy and diagnosis of cancer,
XX      preferably breast cancer, in pharmaceutical compositions, e.g., vaccine
XX      and other compositions for the diagnosis, prevention and treatment of
XX      human malignancies, for stimulating and/or expanding T cells specific for
XX      Her-2/Neu polypeptide and for inhibiting the development of cancer in a
XX      patient. The invention is useful for stimulating a T cell response in a
XX      human patient, as probe or primer for nucleic acid hybridisation, to
XX      selectively form duplex molecules with complementary stretches of the
XX      entire Her-2/Neu gene or gene fragments of interest, to isolate a full
XX      length gene from a suitable library, and to direct expression of a
XX      polypeptide in appropriate host cells. The composition is useful in
XX      prophylactic or therapeutic applications and for the treatment of cancer,
XX      preferably for the immunotherapy of breast cancer and other Her-2/Neu-
XX      associated malignancies. The invention is useful in gene therapy. The
XX      present sequence is human Her-2/neu protein.
SQ      Sequence 1255 AA;
Qy      Query Match
Db      1 MELALCRWGLLALLPPGAASVCTGTDKRLPASPETHLMDRLHLYGQCVQGNL 60
Db      1 MELALCRWGLLALLPPGAASVCTGTDKRLPASPETHLMDRLHLYGQCVQGNL 60
Qy      61 ELTYIPTNASISFLDIOEVQGYVLIANOVROVPLQRLIVRGTLQFEDNYALAVDNG 120
Db      61 ELTYIPTNASISFLDIOEVQGYVLIANOVROVPLQRLIVRGTLQFEDNYALAVDNG 120
Qy      121 DPLNNTTPTVGTASPGGLREIQRSITELIKSGVLIORNPOLCYOQTIIMKQIFHNQOLA 180
Db      121 DPLNNTTPTVGTASPGGLREIQRSITELIKSGVLIORNPOLCYOQTIIMKQIFHNQOLA 180
Qy      181 LTLIDTNSRACHPCSPCKSGRCWGBSSSDQSLRTVCAGGACRCGRLPTDCHEQC 240
Db      181 LTLIDTNSRACHPCSPCKSGRCWGBSSSDQSLRTVCAGGACRCGRLPTDCHEQC 240
Qy      241 AAGCGPRHSDCLCLHNSGICELCPALVYNTTTFEEMPRDEGYTGAACVACP 300
Db      241 AAGCGPRHSDCLCLHNSGICELCPALVYNTTTFEEMPRDEGYTGAACVACP 300
Qy      301 YNYLSTDVSGCTLVCPHNDVETADGTORCEKSKRCARVCYGLGMQYIKANSKFIGIT 360
Db      301 YNYLSTDVSGCTLVCPHNDVETADGTORCEKSKRCARVCYGLGMQYIKANSKFIGIT 360
Qy      361 ELFPAGCKRTIGSLAPFESFDGDPASNTAPLQEQVYFTEETITGYLYSAWPSLP 420
Db      361 ELFPAGCKRTIGSLAPFESFDGDPASNTAPLQEQVYFTEETITGYLYSAWPSLP 420

```

```

Db 361 IOERAGCKIFGSLAFLESFSDGDPASNTAPLOPEQLQVETLEITGYLYISAMPDLSL 420
Qy 421 DLSVFQNTQYIRGRILLNGAYSLTLOGLISWLGSLRSLRSLGLLHHNTLTCVHTV 480
Db 421 DLSVFQNTQYIRGRILLNGAYSLTLOGLISWLGSLRSLRSLGLLHHNTLTCVHTV 480
Qy 481 PMDLFRNPQALHTANRPEDCEVGEGLACHOLCARGHWGSGPIQCVNCSQFLRGQEC 540
Db 481 PMDLFRNPQALHTANRPEDCEVGEGLACHOLCARGHWGSGPIQCVNCSQFLRGQEC 540
Qy 541 VEECRVUQGLPREVYVNRHCLPCHPECOFONGSYTCGPAPADOCVACAHYKDPFCVARC 600
Db 541 VEECRVUQGLPREVYVNRHCLPCHPECOFONGSYTCGPAPADOCVACAHYKDPFCVARC 600
Qy 601 PSQVPLSYMPYIKPFDEBAGACOPCPINCTHSCVDLDGKCAEGRASPLTSIVSAVVG 660
Db 601 PSQVPLSYMPYIKPFDEBAGACOPCPINCTHSCVDLDGKCAEGRASPLTSIVSAVVG 660
Qy 661 ILVVVLGVVFGILIKRQCKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKEFNN 720
Db 661 ILVVVLGVVFGILIKRQCKIRKYMRLLOETELVEPLTPSGAMPNQAQMRILKEFNN 720
Qy 721 FTVSFMLERVPKVSAS-HLETVYKGIWIPDGENVKIPAIKVLBNTPSKANKSLDEAYV 779
Db 720 -----LRKVKVLSGAFGTYYKGIWIPDGENVKIPAIKVLBNTPSKANKSLDEAYV 773
Qy 780 MAGVSPYVSRLLGICLTSTVOLTPMPYGLLDHYRENRGRGLSGDILLNMCQIAKGM 839
Db 774 MAGVSPYVSRLLGICLTSTVOLTPMPYGLLDHYRENRGRGLSGDILLNMCQIAKGM 833
Qy 840 SYLEDVRLVRLDLAARVLYKSPHVKITDPLARLLDDETYVHADGGKVPILKMALES 899
Db 834 SYLEDVRLVRLDLAARVLYKSPHVKITDPLARLLDDETYVHADGGKVPILKMALES 893
Qy 900 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGLPAREIPDLEKGERLPQPICTIDVYM 959
Db 894 ILRRFTHQSDVMSYGVTVWELMTFGAKPYDGLPAREIPDLEKGERLPQPICTIDVYM 953
Qy 960 IMVKCMWIDSECRREFELVSESRNARDQRFVYVIONEDLGPASPLDSTFYSLLEDD 1019
Db 954 IMVKCMWIDSECRREFELVSESRNARDQRFVYVIONEDLGPASPLDSTFYSLLEDD 1013
Qy 1020 MGDLVDAEEYLVPOGFCFPPDPAAGAGMYHRRSSSTRSGGDLTLGSEPSEEARPS 1079
Db 1014 MGDLVDAEEYLVPOGFCFPPDPAAGAGMYHRRSSSTRSGGDLTLGSEPSEEARPS 1073
Qy 1080 PLASBEGASDVFDGDLGMAKAGLQSLPTHDSPIQRYSEDPTVPLPSETDGYVAPLTC 1139
Db 1074 PLASBEGASDVFDGDLGMAKAGLQSLPTHDSPIQRYSEDPTVPLPSETDGYVAPLTC 1133
Qy 1140 SPOREYVNOPDVRQPPSPREGPLPARPAGATLERAKTLSPGNGVYKVFAPGAVEN 1199
Db 1134 SPOREYVNOPDVRQPPSPREGPLPARPAGATLERAKTLSPGNGVYKVFAPGAVEN 1193
Qy 1200 PEVYLTPOGGAAPQHPAPSPAFDNLVYWDQPEPAGPSTFKGTPTANPEVYGLDV 1259
Db 1194 PEVYLTPOGGAAPQHPAPSPAFDNLVYWDQPEPAGPSTFKGTPTANPEVYGLDV 1253
Qy 1260 PV 1261
Db 1254 PV 1255

RESULT 13
AAMS1143
ID AAMS1143 standard; Protein: 1255 AA.
XX
AC AAMS1143;
XX
DT 17-JUN-2002 (first entry)
XX
DE Human Her-2/neu oncogene-encoded p185 glycoprotein.
XX

```

```

KW Her-2/neu; oncogene; cancer; tumour; vaccine; human; p185;
KM tyrosine kinase; receptor; c-erbB2; gene therapy.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 1..653
XX FT Domain /note= "extracellular domain"
XX FT Domain 676..1255
XX FT Domain /note= "intracellular domain"
XX FT Domain 990..1255
XX FT /note= "phosphorylation domain"
XX
XX MO200212341-A2.
XX
XX 14-FEB-2002.
XX
XX 03-AUG-2001; 2001MO-US24283.
XX
XX 03-AUG-2000; 2000US-0632507.
XX
XX (CORI-) CORIXA CORP.
XX (SMIK) SMTHTKLINB BEECHAM BIOLOGICALS.
XX
XX Cheever MA, Gheysen D;
XX
XX WPI; 2002-241743/29.
XX DR N-PSDB; ABA92250.
XX
XX Her-2/neu fusion protein for treating or preventing cancer by eliciting
XX PT or enhancing an immune response to the protein, has Her-2/neu
XX PT extracellular domain fused to Her-2/neu intracellular or
XX PT phosphorylation domain
XX
XX Claim 68; Fig 7; 141pp; English.
XX
XX The present sequence is that of human Her-2/neu (p185 glycoprotein
XX or c-erbB2), an oncogenic self-protein and target for anti-cancer
XX vaccines. The Her-2/neu gene is amplified and p185 is overexpressed
XX in a variety of cancers, including breast, ovarian, colon, lung and
XX prostate cancer. Her-2/neu is a member of the tyrosine kinase
XX family of receptor-like glycoproteins. It comprises an extracellular
XX domain with homology to the epidermal growth factor receptor
XX (EGFR), a highly hydrophobic transmembrane domain and a C-terminal
XX intracellular domain that also shows homology to RGR. Its
XX overexpression correlates with a poor prognosis in breast and
XX ovarian cancers. The invention provides Her-2/neu fusion
XX proteins, nucleic acids encoding them, viral vectors, and vaccines
XX comprising the fusion proteins or nucleic acid molecules. In
XX preferred fusion proteins, the extracellular domain of a Her-2/neu
XX protein is fused to a Her-2/neu intracellular domain or
XX phosphorylation domain (or its DeltaCD fragment). An immune
XX response to Her-2/neu protein is elicited or enhanced by
XX administering the fusion protein in the form of a vaccine, or by
XX transfecting cells of an animal ex vivo with a nucleic acid
XX encoding the fusion protein, and delivering the transfected cells
XX to the animal. The fusion proteins, nucleic acids, and isolated
XX specific T-cells are useful for inhibiting the development of a
XX cancer, especially breast, ovarian, colon, lung or prostate cancer
XX in a patient. T cells that specifically react with a Her-2/neu
XX fusion protein can be used to remove tumour cells from a sample in
XX order to inhibit the development of cancer in a patient.
XX
XX Sequence 1255 AA;
XX
XX Query Match 96.9%; Score 6641; DB 23; Length 1255;
XX Best Local Similarity 97.3%; Pred. No. 0;
XX Matches 1228; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
Qy 1 METALCRWGLLALPLPGAASTGYCTGDKMLRPPAPETHLDLRLYGGCQVYQGNL 60
Db 1 METALCRWGLLALPLPGAASTGYCTGDKMLRPPAPETHLDLRLYGGCQVYQGNL 60

```

```

QY 61 ELTYLPTNASLSFQDIOEVGVYLIANQVRYVLRIRIVRGTOLEFEDNYALAVLDNG 120
Db 61 ELTYLPTNASLSFQDIOEVGVYLIANQVRYVLRIRIVRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTPTVTSASFGSLRELOLRSLTEILKGVLIQGNPOLCYODTILMKDIFKNNOLA 180
Db 121 DPLNNTPTVTSASFGSLRELOLRSLTEILKGVLIQGNPOLCYODTILMKDIFKNNOLA 180
QY 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTFTVACGACARCKGLPTDCHEQC 240
Db 181 LTLIDTNRSRACHPCSPCKSGRCWGESSEDCQSLTFTVACGACARCKGLPTDCHEQC 240
QY 241 AAGCTGPRGSDCLCLPHNSGICELHCPALVTNTOTPEMNPBERGYTFGASCTACP 300
Db 241 AAGCTGPRGSDCLCLPHNSGICELHCPALVTNTOTPEMNPBERGYTFGASCTACP 300
QY 301 YNYLSTDVGSCTLVCPHNEVTAEDGTORCEKSKPCARVCYGLAGQYIKANSKFIGIT 360
Db 301 YNYLSTDVGSCTLVCPHNEVTAEDGTORCEKSKPCARVCYGLAGQYIKANSKFIGIT 360
QY 361 ELFFAGCKKIFGSLAFLEPSFDGDPASTAPLOEQVPELTLEITGYLYISAMPDLP 420
Db 361 ELFFAGCKKIFGSLAFLEPSFDGDPASTAPLOEQVPELTLEITGYLYISAMPDLP 420
QY 421 DLSVFQNLQVIRGRILHNGAYSITLQGLISWLGSLREISGIALIHHNTHLCFVHTV 480
Db 421 DLSVFQNLQVIRGRILHNGAYSITLQGLISWLGSLREISGIALIHHNTHLCFVHTV 480
QY 481 PMDOLFRNPHOALLHTANRPDECEVGBGLACHOLCAFGHCMPGPTQCVNCSQFLRQEC 540
Db 481 PMDOLFRNPHOALLHTANRPDECEVGBGLACHOLCAFGHCMPGPTQCVNCSQFLRQEC 540
QY 541 VEEGRVLOGLPREVYNNARHCLPCHPECOPOKGSVTCGPEADQCAAHYKDPFCVARC 600
Db 541 VEEGRVLOGLPREVYNNARHCLPCHPECOPOKGSVTCGPEADQCAAHYKDPFCVARC 600
QY 601 PSYKPDLSYMPIMKPFDEBACQPCPINCTHSCVDLDKCPAEQASPLTISAVVG 660
Db 601 PSYKPDLSYMPIMKPFDEBACQPCPINCTHSCVDLDKCPAEQASPLTISAVVG 660
QY 661 ILVVVILGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSAMNQOMRLKEFNN 720
Db 661 ILVVVILGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSAMNQOMRLKEFNN 720
QY 721 FTVSFMLRVPKVSAS-HLETVYKGIWIPDGENVKIPAIKYLRENTSPKANKETLDEAYV 779
Db 721 FTVSFMLRVPKVSAS-HLETVYKGIWIPDGENVKIPAIKYLRENTSPKANKETLDEAYV 779
QY 779 MAGVGSPPYVSRLLGICLTSTVQVLTQMLPYGCLLDHYENRGRIGSODLLINMCQIAKGM 833
Db 779 MAGVGSPPYVSRLLGICLTSTVQVLTQMLPYGCLLDHYENRGRIGSODLLINMCQIAKGM 833
QY 833 SYLEDVLRHDLARAVLVKSPYHVKITDGLARLIDIDETEHADGGVPIKMMLES 899
Db 833 SYLEDVLRHDLARAVLVKSPYHVKITDGLARLIDIDETEHADGGVPIKMMLES 899
QY 899 ILRRRFTHQSDVMSYGVYTWELMTFGAKPYDGPAREIPDLEKGERLPORPCTIDVYM 959
Db 899 ILRRRFTHQSDVMSYGVYTWELMTFGAKPYDGPAREIPDLEKGERLPORPCTIDVYM 959
QY 959 IMVCKMWIDSECRREFELVSEFSRMARDPORFVYIONEDLGPASPLDSTFYSLLEDD 1013
Db 959 IMVCKMWIDSECRREFELVSEFSRMARDPORFVYIONEDLGPASPLDSTFYSLLEDD 1013
QY 1013 MGDVDAEYLVPCQGFCEPDPAAGAGVYHRRHSSSTRSGGDLTLGLEPSEBEAPRS 1079
Db 1013 MGDVDAEYLVPCQGFCEPDPAAGAGVYHRRHSSSTRSGGDLTLGLEPSEBEAPRS 1079
QY 1079 PLAPSEBAGSDVPDGLGMAAKGLQSLPTHDSPLQRYSEDPVPLPSTTDGVVAPLTC 1139
Db 1079 PLAPSEBAGSDVPDGLGMAAKGLQSLPTHDSPLQRYSEDPVPLPSTTDGVVAPLTC 1139
QY 1139 SPOEYVNOFDVBPQSPREGLPARPAGATLERAKTLPNGNIVXDVAFGAVEN 1199

```

```

Db 1134 SPOEYVNOFDVBPQSPREGLPARPAGATLERAKTLPNGNIVXDVAFGAVEN 1199
QY 1200 PEYLTPQGAAPQHPHPAFDFLYWDDPPERGAPESTFKGTPTAENPEYGLDV 1259
Db 1194 PEYLTPQGAAPQHPHPAFDFLYWDDPPERGAPESTFKGTPTAENPEYGLDV 1253
QY 1260 PV 1261
Db 1254 PV 1255

RESULT 14
AAU77114
ID AAU77114 standard; Protein, 1255 AA.
XX
AC AAU77114;
XX
DT 05-JUN-2002 (first entry)
XX
DE Human Her-2/neu polypeptide.
XX
KW Human; Her-2/neu; cytosolic; haematological malignancy; CML;
KW acute myelogenous leukaemia; AML; chronic myelogenous leukaemia; CLL;
KW chronic lymphocytic leukaemia; myeloma; non-Hodgkin's lymphoma; MDS;
KW Hodgkin's lymphoma; T cell therapy.
XX
OS Homo sapiens.
XX
PN WO200213847-A2.
XX
PD 21-FEB-2002.
XX
PF 13-AUG-2001; 2001MO-US25408.
XX
PR 14-AUG-2000; 2000US-0638280.
XX
PR 28-SEP-2000; 2000US-0675904.
XX
PA (CORI-) CORIXA CORP.
XX
PI Gaiger A, Cheever MA, Hand-zimmermann S,
XX
DR N-PSDB; ABK10730.
XX
DR N-PSDB; ABK10730.
XX
PT Inhibiting haematological malignancy development by administering
PT polypeptide comprising immunogenic portion of Her-2/neu, polynucleotide
PT encoding the polypeptide, or antigen presenting cells expressing the
PT polypeptide
XX
PS Disclosure; Page 71-74; 74pp; English.
XX
CC The invention relates to a method for inhibiting development of
CC haematological malignancy in a patient by administering a polypeptide
CC encoding an immunogenic portion of Her-2/neu or a polynucleotide
CC encoding the polypeptide. Antigen presenting cells that express the
CC protein can also be administered. The sequences are used for inhibiting
CC development of haematological malignancy such as acute myelogenous
CC leukaemia (AML), chronic myelogenous leukaemia (CML), chronic lymphocytic
CC leukaemia (CLL), MDS, myelomas, Hodgkin's lymphoma and non-Hodgkin's
CC lymphoma. This sequence represents the human Her-2/neu polypeptide.
XX
SQ Sequence 1255 AA;
XX
Query Match 96.9%; Score 6641; DB 23; Length 1255;
Best Local Similarity 97.3%; Pred. No. 0;
Matches 1288; Conservative 5; Mismatches 21; Indels 8; Gaps 2;
QY 1 MELALCRWGLLIALPPGAASYCTGTDMKLRIPASPTHLDMLRHLYQCGQVVGNTL 60
Db 1 MELALCRWGLLIALPPGAASYCTGTDMKLRIPASPTHLDMLRHLYQCGQVVGNTL 60
QY 61 ELTYLPTNASLSFQDIOEVGVYLIANQVRYVLRIRIVRGTOLEFEDNYALAVLDNG 120

```

```

61 ELTYLPTNASTLSTFLOIOEVGVYLIANQVROVPLQRLIVRTQGLFEDNYALAVLDNG 120
121 DPLNNTTPVTGASPGGLRELOJRSLEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
121 DPLNNTTPVTGASPGGLRELOJRSLEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
181 LTLIDTNSRACHPCSPCKGSRCKGSESEDCOSLTRIVCAGGACRCKGPLPTDCHEQC 240
181 LTLIDTNSRACHPCSPCKGSRCKGSESEDCOSLTRIVCAGGACRCKGPLPTDCHEQC 240
241 AAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
241 AAGCTGPRGSDCLACLFHNSGICELHCPALVTYNTDTFESMPNPEGRTYFGASCVTACP 300
301 YNYLSTDVSGCTIVCPLNCEVTAEDGTORCEKSKPCARVCYGLGSHLEVRVATVSAN 360
301 YNYLSTDVSGCTIVCPLNCEVTAEDGTORCEKSKPCARVCYGLGSHLEVRVATVSAN 360
361 ELLEFAGCKKIFGSLAFLEPESFDGDPASNTAPLOPEOLQVETLEITGYLYISAMPDLP 420
361 IOEFAGCKKIFGSLAFLEPESFDGDPASNTAPLOPEOLQVETLEITGYLYISAMPDLP 420
421 DLSVFQNLQVIRGRILNHGASLTLOGLISMLGRSLRELSGLALIHNTHLCFVHTV 480
421 DLSVFQNLQVIRGRILNHGASLTLOGLISMLGRSLRELSGLALIHNTHLCFVHTV 480
481 PMDOLFRNPHQALHTANRPEDECEVGEGLACHOLCARHCWGPPTOCVNCOSQFLRQEC 540
481 PMDOLFRNPHQALHTANRPEDECEVGEGLACHOLCARHCWGPPTOCVNCOSQFLRQEC 540
541 VEECKVLOGLFREYVNAHCLPCHPECOPONGSYTCGPEADQCVACAHYDPPFCVACR 600
541 VEECKVLOGLFREYVNAHCLPCHPECOPONGSYTCGPEADQCVACAHYDPPFCVACR 600
601 PSQVPLDSTYMPIMKFPDEGACQPCPINCCHSCVDLDDKCPREAPSPILTSIISAVG 660
601 PSQVPLDSTYMPIMKFPDEGACQPCPINCCHSCVDLDDKCPREAPSPILTSIISAVG 660
661 ILVVVIGVFGIILKRRQOKIRKYMRLLOETELVEPLTPSGAMPNOQMRLKEFN 720
661 ILVVVIGVFGIILKRRQOKIRKYMRLLOETELVEPLTPSGAMPNOQMRLKEFN 720
721 FTVSFMLEVPRVNAS-HLETVYKGIWIPDGENVNIPIVAKLRENTSKAKETLDEAYV 779
721 FTVSFMLEVPRVNAS-HLETVYKGIWIPDGENVNIPIVAKLRENTSKAKETLDEAYV 779
780 MAGVSPYVSRLLGLCLTSTVQVLTQMLPYGCLLDHYRENRGLSGODLNTMCMQIAKGM 839
780 MAGVSPYVSRLLGLCLTSTVQVLTQMLPYGCLLDHYRENRGLSGODLNTMCMQIAKGM 839
834 SYLEPVLVHRDLAARNLVKSPNHVKITDFGLARLIDIDETVHADGKVPITWMALES 899
834 SYLEPVLVHRDLAARNLVKSPNHVKITDFGLARLIDIDETVHADGKVPITWMALES 899
894 ILRRFTHQSDVMSYGVTVMEIMTFGAKPYDGPAREIPDLLEKGERLPQPCITIVYM 959
894 ILRRFTHQSDVMSYGVTVMEIMTFGAKPYDGPAREIPDLLEKGERLPQPCITIVYM 959
954 IMVKCMWIDSECRPFRELVSEFSMARDPQRFVVIQNEDELGPASPJLSTYRSLDDEDD 1019
954 IMVKCMWIDSECRPFRELVSEFSMARDPQRFVVIQNEDELGPASPJLSTYRSLDDEDD 1019
1020 MODLVDAEELVPOCGFPCPDPAAGAGMVAHRRSSSTRSGGDLTLGLEPSEBEAPRS 1079
1020 MODLVDAEELVPOCGFPCPDPAAGAGMVAHRRSSSTRSGGDLTLGLEPSEBEAPRS 1079
1074 FLAPSEGASDVFDGDLGMAKAGISLPTHDSPFLQYSEDFTVPLPSETDGYVAPLTC 1139
1074 FLAPSEGASDVFDGDLGMAKAGISLPTHDSPFLQYSEDFTVPLPSETDGYVAPLTC 1139
1140 SPOPEYVNOQDVPRPQPPSPREGPLPAAPAGATLERAKTSLPGKNGVQVDFAFGAVEN 1199
1140 SPOPEYVNOQDVPRPQPPSPREGPLPAAPAGATLERAKTSLPGKNGVQVDFAFGAVEN 1199

```

```

Db 1134 SPOPEYVNOQDVPRPQPPSPREGPLPAAPAGATLERAKTSLPGKNGVQVDFAFGAVEN 1193
Qy 1200 PEYLTGGAAPQPPPPAFSPAFNLYVWDDPPBERGAPSTFGTPTAENPEYLGUDV 1259
Db 1194 PEYLTGGAAPQPPPPAFSPAFNLYVWDDPPBERGAPSTFGTPTAENPEYLGUDV 1253
Qy 1260 PV 1261
Db 1254 PV 1255

RESULT 15
AAR39568
ID AAR39568 standard; Protein; 1433 AA.
XX
AC AAR39568;
XX
DT 07-FEB-1994 (first entry)
XX
DE Sequence of c-erbB-2 tumour antigen.
XX
KW Tumour antigen; c-erbB-2; glycoprotein.
XX
OS Homo sapiens.
XX
PN MOJ316185-A.
XX
PD 19-AUG-1993.
XX
PF 05-FEB-1993; 93MO-US01055.
XX
PR 06-FEB-1992; 92US-0831967.
XX
PA (CETU) CETUS ONCOLOGY CORP.
PA (CREA-) CREATIVE BIOMOLECULES INC.
XX
PI Houston LL, Huston JS, Oppermann H, Ring DB;
DR WPI; 1993-272889/34.
DR N-FSDB; AAQ46083.
XX
PT New single chain Fv polypeptide binding to C-erbB-2 tumour
PT antigen - for imaging or treating breast or ovarian cancer etc.
XX
PS Disclosure; pages 48-54; 87pp; English.
XX
CC c-erbB-2 refers to a protein antigen expressed on the surface of
CC tumour cells, such as breast and ovarian tumour cells, which is an
CC approx. 200,000 mol. wt. acidic glycoprotein having an isoelectric
CC pt. of about 5.3 (see AAQ46083, AAR39568). X in AAR39563 represents
CC the location of a stop codon in AAQ46083.
XX
SQ Sequence 1433 AA;

Query Match 96.3%; Score 6598; DB 14; Length 1433;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1221; Conservative 7; Mismatches 26; Indels 8; Gaps 2;

Qy 1 MELALCRWGLIALILPPGAASTOVCTGDMKRLRPASPEHLDMRLHYOVQVQGNL 60
Db 1 MELALCRWGLIALILPPGAASTOVCTGDMKRLRPASPEHLDMRLHYOVQVQGNL 60
Qy 61 ELTYLPTNASTLSTFLOIOEVGVYLIANQVROVPLQRLIVRTQGLFEDNYALAVLDNG 120
Db 61 ELTYLPTNASTLSTFLOIOEVGVYLIANQVROVPLQRLIVRTQGLFEDNYALAVLDNG 120
Qy 121 DPLNNTTPVTGASPGGLRELOJRSLEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Db 121 DPLNNTTPVTGASPGGLRELOJRSLEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
Qy 181 LTLIDTNSRACHPCSPCKGSRCKGSESEDCOSLTRIVCAGGACRCKGPLPTDCHEQC 240
Db 181 LTLIDTNSRACHPCSPCKGSRCKGSESEDCOSLTRIVCAGGACRCKGPLPTDCHEQC 240

```

Search completed: July 22, 2003, 09:16:49  
 Job time : 42.1104 secs

```

QY 241 AACCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTESMENPEGRYTFGASCTYACP 300
Db 241 AACCTGPKHSDCLACLFHNSGICELHCPALVTYNTDTESMENPEGRYTFGASCTYACP 300
QY 301 YNLTSTDVSGCTLVCPAHNOEVTAEADTORCEKSKPCARVCYGLMOYIKANSKFIGIT 360
Db 301 YNLTSTDVSGCTLVCPAHNOEVTAEADTORCEKSKPCARVCYGLMEHLREYRAVTSAN 360
QY 361 ELEFFAGCKKIFGSLAFLPESFPDGPASNTAPLQPEQLOVETLEETITGYLYISAMPDLP 420
Db 361 IQEFAGCKRIFGSLAFLPESFPDGPASNTAPLQPEHLOVETLEETITGYLYISAMPDLP 420
QY 421 DLSVFQNLQYIRGRILHNGAYSLTQGLGISMLGLSLRELSGLALIHNTILCFVATV 480
Db 421 DLSVFQNLQYIRGRILHNGAYSLTQGLGISMLGLSLRELSGLALIHNTILSFVATV 480
QY 481 PMQOLFRNPHQALHTANRPEDCEVGEGLACHQLCARGHMGSPFTQCVNCSQFLRGQEC 540
Db 481 PMQOLFRNPHQALHTANRPEDCEVGEGLACHQLCARGHMGSPFTQCVNCSQFLRGQEC 540
QY 541 VEECRVILQGLPREYVNAHCLPCHPECOFONGSVTCGPBADQCACAHYKDPFCVAC 600
Db 541 VEECRVILQGLPREYVNAHCLPCHPECOFONGSVTCGPBADQCACAHYKDPFCVAC 600
QY 601 PSGVKPDLASYMPIWKPFDEEGACOPINCTHSCVDLDDKCPAEQASPLTSIVSAVVG 660
Db 601 PSGVKPDLASYMPIWKPFDEEGACOPINCTHSCVDLDDKCPAEQASPLTSIVSAVVG 660
QY 661 ILLVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOAOMRILKEFNN 720
Db 661 ILLVVLGVVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOAOMRILKEFNN 720
QY 721 FTVSFMLRVKVSAS-HLETVYKGIWIPDGENVKIPALIVLRENTSPKANKELDEAYV 779
Db 721 FTVSFMLRVKVSAS-HLETVYKGIWIPDGENVKIPALIVLRENTSPKANKELDEAYV 779
QY 780 MAGVSPYVSRLIGICTSTVQLVTOCLMPYGCLLDHYRENRGRLSGCDLLNMCMQIAKGM 839
Db 780 MAGVSPYVSRLIGICTSTVQLVTOCLMPYGCLLDHYRENRGRLSGCDLLNMCMQIAKGM 839
QY 840 SYLEDVRLVHRLAARNVLKSPNHVKITDFGLARLLIDETEHADGSKVPIKMALES 899
Db 840 SYLEDVRLVHRLAARNVLKSPNHVKITDFGLARLLIDETEHADGSKVPIKMALES 899
QY 900 ILRRRFTHOSDVMSTGYTWELMTFGAKPYDGI PAREIPDLEKGERLPQPPICITDVM 959
Db 900 ILRRRFTHOSDVMSTGYTWELMTFGAKPYDGI PAREIPDLEKGERLPQPPICITDVM 959
QY 954 IMVKCWMIDSECRPRRELVSFSSMARADPQRFVIONEDLGPASPLDSTFYASLLEDD 1019
Db 954 IMVKCWMIDSECRPRRELVSFSSMARADPQRFVIONEDLGPASPLDSTFYASLLEDD 1019
QY 1020 MGLVDAEEYLVPOQGFCCPDPAFGAGMHHRRSSSTRSGGDLTLGLEPSEEPAPRS 1079
Db 1020 MGLVDAEEYLVPOQGFCCPDPAFGAGMHHRRSSSTRSGGDLTLGLEPSEEPAPRS 1079
QY 1080 PLAPSEGAGSDVDGLGMAAKGLOSLPTHDSPLQRYSEDPVPLPSETDGYVAPLTC 1139
Db 1080 PLAPSEGAGSDVDGLGMAAKGLOSLPTHDSPLQRYSEDPVPLPSETDGYVAPLTC 1139
QY 1140 SPQPEYVNDPVPQPPSPREGPLPAARPAATLERAKTILSPKNGVAVKVPFAGAVEN 1199
Db 1140 SPQPEYVNDPVPQPPSPREGPLPAARPAATLERAKTILSPKNGVAVKVPFAGAVEN 1199
QY 1200 PEYLTPQGAAPQPHPPAPSPAFNDLYYWDODPPERGAAPPSTFKGPTAENPEYLGLDV 1259
Db 1200 PEYLTPQGAAPQPHPPAPSPAFNDLYYWDODPPERGAAPPSTFKGPTAENPEYLGLDV 1259
QY 1260 PV 1261
Db 1260 PV 1261
QY 1254 PV 1255
Db 1254 PV 1255

```



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:10:23 ; Search time 22.2855 Seconds

(without alignments)  
5413.772 Million cell updates/sec

Title: SEQ4-710-730-14

Sequence: 1 MELALCRWGLLALLPPCA.....TFKGTPTANPEYLGIDUV 1255

Scoring table: BIOSOM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Fred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6602	96.9	1255	1 A24571	protein-tyrosine k
2	5805	85.2	1260	1 TVRNU	protein-tyrosine k
3	5795.5	85.1	1254	2 I48161	p-185 precursor
4	3035	44.5	1210	1 GQHB	epidermal growth f
5	3006	44.1	1210	2 AS183	epidermal growth f
6	2983.5	43.8	1223	1 TVCHLV	epidermal growth f
7	2852.5	41.9	1308	2 A47253	epidermal growth f
8	2559	37.6	1166	1 S06142	protein-tyrosine k
9	2307.5	33.9	1342	2 A36223	kinase-related tra
10	2321.5	32.6	1339	2 TC4827	epidermal growth f
11	1655.5	24.3	688	1 TVPLV	epidermal growth f
12	1592	23.4	604	1 TVYOH	protein-tyrosine k
13	1541	22.6	544	2 S35745	protein-tyrosine k
14	1534	22.5	545	2 S00727	protein-tyrosine k
15	1517	22.3	540	2 B44776	kinase-related tra
16	1515	22.2	540	1 TVPEB	protein-tyrosine k
17	1514.5	22.2	1330	1 GQFPE	epidermal growth f
18	1509	22.1	644	2 A36125	epidermal growth f
19	1217	17.9	1373	2 E88257	protein-tyrosine k
20	1217	17.9	1374	2 S70712	protein-tyrosine k
21	1131	16.6	1369	2 S70713	protein-tyrosine k
22	1126	16.5	527	2 A42032	epidermal growth f
23	1114	16.4	1717	1 A45558	epidermal growth f
24	975.5	14.3	843	2 A27331	epidermal growth f
25	806.5	11.8	346	2 S13807	epidermal growth f
26	754.5	11.1	311	2 S13808	protein-tyrosine k
27	693	10.2	1363	2 T43220	insulin-like growt
28	657	9.6	1382	1 INHUR	insulin receptor p
29	656.5	9.6	1372	2 A34157	insulin receptor p

30	653	9.6	1383	2 A36080	insulin receptor p
31	643.5	9.4	1300	2 A36502	insulin receptor-r
32	632	9.3	1477	2 T18534	protein-tyrosine k
33	630	9.2	1607	2 T43212	insulin-like growt
34	629	9.2	1268	2 B35502	insulin receptor-r
35	602	8.8	1367	1 TGHUR1	insulin-like growt
36	583	8.6	1371	2 A33837	insulin-like growt
37	581.5	8.5	2148	1 A56081	insulin receptor
38	580	8.5	1114	1 S05582	protein-tyrosine k
39	579	8.5	987	2 A54092	protein-tyrosine k
40	575	8.4	2101	2 S57245	insulin receptor
41	569	8.4	1390	2 T30346	insulin receptor
42	559	8.2	987	2 I48652	mouse developmenta
43	551.5	8.1	952	2 I50612	protein-tyrosine k
44	548.5	8.1	801	4 TVHURE	transforming prote
45	548.5	8.1	984	2 A39753	protein-tyrosine k

## ALIGNMENTS

### RESULT 1

A24571

protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human

N/Alternate names: c-erb-B-2 protein precursor; kinase-related transforming protein

C/Species: Homo sapiens (man)

C/Date: 25-Oct-1987 #sequence revision 06-Dec-1996 #text change 11-Jun-1999

C/Accession: A24571; A25491; A44188; B44188; I59509; I57622

R/Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, Nature 319, 230-234, 1986

A/Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal grow

A/Reference number: A24571; MUID:86118663; PMID:3003577

A/Accession: A24571

A/Molecule type: mRNA

A/Residues: 1-1255 <YAM>

A/Cross-references: GB:X0363; NID:G31197; PIDN:CAA27060.1; PID:G31198

R/Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.

Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985

A/Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epi

A/Reference number: A25491; MUID:86016729; PMID:2995967

A/Accession: A25491

A/Molecule type: DNA

A/Residues: 737-1031 <SEM>

A/Cross-references: GB:M1767; NID:G182163; PIDN:AAA35808.1; PID:G553282

R/Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg Science 230, 1132-1139, 1985

A/Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chr

A/Reference number: A44188; MUID:86070181; PMID:2999974

A/Accession: A44188

A/Molecule type: DNA

A/Residues: 740-910 <COU>

A/Cross-references: GB:M12036; NID:G183988; PIDN:AAA35978.1; PID:G183989

A/Accession: B44188

A/Molecule type: mRNA

A/Residues: 1-517; RALU, 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>

R/King, C.R.; Kraus, M.H.; Aaronson, S.A.

Science 229, 974-976, 1985

A/Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.

A/Reference number: I59509; MUID:8527597; PMID:2992089

A/Accession: I59509

A/Molecule type: DNA

A/Status: translated from GB/EMBL/DBJ

A/Residues: 832-909 <REX>

A/Cross-references: GB:L29395; NID:G459807; PIDN:AAA35809.1; PID:G459808

R/Rat, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.

Mol. Cell. Biol. 7, 2597-2601, 1987

A/Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio

A/Reference number: I57622; MUID:87268698; PMID:3039351

A/Accession: I57622

A/Molecule type: DNA

A/Status: translated from GB/EMBL/DBJ

A/Residues: 1-191 <TAU>

A:Cross-references: GB:M16792; NID:9183983; PIDN:AAA58637.1; PID:9553332  
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30  
 C:Genetics:  
 A:Gene: GDB:ERBB2; NCL; NEU; HER-2  
 A:Cross-references: GDB:120613; OMTM:164870  
 A:Map position: 17q21.1-17q21.1  
 A:Introns: 25/1; 75/3; 147/1; 883/3  
 A>Note: the list of introns is incomplete  
 C:Function:  
 A:Description: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 inase  
 F:1-21/Domain: signal sequence #status predicted <SIG>  
 F:12-125/Domain: protein-tyrosine kinase erbB2 #status predicted <MAT>  
 F:12-653/Domain: extracellular #status predicted <EXT>  
 F:170-304/Domain: EGF receptor extracellular domain repeat <EP1>  
 F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>  
 F:654-675/Domain: transmembrane #status predicted <TM>  
 F:676-1255/Domain: intracellular #status predicted <INT>  
 F:716-983/Domain: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif  
 F:686-124,187,229,530,571,629/Binding site: carbohydrate (asn) (covalent) #status predicted  
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:753/Active site: Lys #status predicted  
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 96.9%; Score 6602; DB 1; Length 1255;  
 Best Local Similarity 96.9%; Pred. No. 4.6e-277;  
 Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

```

OY 1 MELALCGWGLLLLPAGASTQVCTGDKMLRLPASPETHLMDLRLHLYGCGVQGNL 60
DB 1 MELALCGWGLLLLPAGASTQVCTGDKMLRLPASPETHLMDLRLHLYGCGVQGNL 60
OY 61 ELTYLPTNASLSPLODIOEVGYVLIANQVQVPLRLRIRAGTQLFEDNYALAVDNG 120
DB 61 ELTYLPTNASLSPLODIOEVGYVLIANQVQVPLRLRIRAGTQLFEDNYALAVDNG 120
OY 121 DPLNNTTPVYTCASGGLRELOLRSLETLKGVLIQGNPOLCYDITLKMIDPHKNNOLA 180
DB 121 DPLNNTTPVYTCASGGLRELOLRSLETLKGVLIQGNPOLCYDITLKMIDPHKNNOLA 180
OY 121 DPLNNTTPVYTCASGGLRELOLRSLETLKGVLIQGNPOLCYDITLKMIDPHKNNOLA 180
DB 121 DPLNNTTPVYTCASGGLRELOLRSLETLKGVLIQGNPOLCYDITLKMIDPHKNNOLA 180
OY 181 LTLIDNRSRACHPCSPKCKSRGCGSSSDCOSLTRVCGGACRCKGPLPDDCCEOC 240
DB 181 LTLIDNRSRACHPCSPKCKSRGCGSSSDCOSLTRVCGGACRCKGPLPDDCCEOC 240
OY 241 AAGCTGPRHSDCLALHFNHSGICELHCPALVTYNTDTFESMPNBSGRYTFGASCVTACP 300
DB 241 AAGCTGPRHSDCLALHFNHSGICELHCPALVTYNTDTFESMPNBSGRYTFGASCVTACP 300
OY 301 YNYISTDVGSCTLVCPHNOEVTADGTCRCKSKCARCYVGLGMQYIKANSKFIGIT 360
DB 301 YNYISTDVGSCTLVCPHNOEVTADGTCRCKSKCARCYVGLGMQYIKANSKFIGIT 360
OY 301 YNYISTDVGSCTLVCPHNOEVTADGTCRCKSKCARCYVGLGMQYIKANSKFIGIT 360
DB 301 YNYISTDVGSCTLVCPHNOEVTADGTCRCKSKCARCYVGLGMQYIKANSKFIGIT 360
OY 361 ELLEFAGCKKIFGSLAFPESEFGDPASNTAPLQEOQLVEFTLEITGYLYISAMPDLP 420
DB 361 ELLEFAGCKKIFGSLAFPESEFGDPASNTAPLQEOQLVEFTLEITGYLYISAMPDLP 420
OY 361 IOEFAGCKKIFGSLAFPESEFGDPASNTAPLQEOQLVEFTLEITGYLYISAMPDLP 420
DB 361 IOEFAGCKKIFGSLAFPESEFGDPASNTAPLQEOQLVEFTLEITGYLYISAMPDLP 420
OY 421 DLSTFQNLQVIRGSLIHNGAYSLTLOGISWILRLRLRELGSLAIHNNTHLCFHTYV 480
DB 421 DLSTFQNLQVIRGSLIHNGAYSLTLOGISWILRLRLRELGSLAIHNNTHLCFHTYV 480
OY 481 PMDOLFRNFQALHTANRPEDECVEGLACHQLCARAGHGWGPPTOCVAVCSQRLRQEC 540
DB 481 PMDOLFRNFQALHTANRPEDECVEGLACHQLCARAGHGWGPPTOCVAVCSQRLRQEC 540
OY 541 VEEGRVYQGLPREVYNNARHCLPCHPECOQNGSTTCGPRADQVAAAHYKDPFCYARC 600
DB 541 VEEGRVYQGLPREVYNNARHCLPCHPECOQNGSTTCGPRADQVAAAHYKDPFCYARC 600
OY 601 PSQVKKPDLSTYPIWKFPDEBAGACPCPCINCTHSCVDLDDGCPAEQASPLTSTVSAVVG 660
DB 601 PSQVKKPDLSTYPIWKFPDEBAGACPCPCINCTHSCVDLDDGCPAEQASPLTSTVSAVVG 660

```

```

DB 601 PSQVKKPDLSTYPIWKFPDEBAGACPCPCINCTHSCVDLDDGCPAEQASPLTSTVSAVVG 660
OY 661 ILLVVLGVVFGILLIRROOKIRKYMRLLOETLVEPLTPSGAMPNOAQRILLKETEL 720
DB 661 ILLVVLGVVFGILLIRROOKIRKYMRLLOETLVEPLTPSGAMPNOAQRILLKETEL 720
OY 721 RKVKVLGSAFGFNNFTVSFWLRVP----KVSASHLEVRNRENTSPKANKEILDEAYVMA 775
DB 721 RKVKVLGSAFGF---TVYKGIWIDGENVKIPVA-IKVLRNRENTSPKANKEILDEAYVMA 775
OY 776 GVGSPVYSVLLIGLSTSTVQLTQMLPYCCLLDHRENRGRSGDILLNMCQIAKMSY 835
DB 776 GVGSPVYSVLLIGLSTSTVQLTQMLPYCCLLDHRENRGRSGDILLNMCQIAKMSY 835
OY 836 LEDVRVHVDLAARVNLVSPHNVKITDGLARLLDIBETVHAQDGKPIKMALESL 895
DB 836 LEDVRVHVDLAARVNLVSPHNVKITDGLARLLDIBETVHAQDGKPIKMALESL 895
OY 896 RRRFTHQSDVMSYGVTTWELMTFGAKPYDGIAREIPDLLEKGERLPQPICTIDVYIM 955
DB 896 RRRFTHQSDVMSYGVTTWELMTFGAKPYDGIAREIPDLLEKGERLPQPICTIDVYIM 955
OY 956 VKCWMIDSCRPFRFELVSESRMARDPQRFVIONEDLGPASPLDSTFYRLLEDDMG 1015
DB 956 VKCWMIDSCRPFRFELVSESRMARDPQRFVIONEDLGPASPLDSTFYRLLEDDMG 1015
OY 1016 DIVDAEYLVPOQGFPCDPAPAGAGVHRRSSSTRSGGDLTLGLEPSEEAAPRSL 1075
DB 1016 DIVDAEYLVPOQGFPCDPAPAGAGVHRRSSSTRSGGDLTLGLEPSEEAAPRSL 1075
OY 1076 ABSEAGSVDFPDGLGMGAKGLQSLPTHDPSPRLRYSDDPVLPLSEEDGVAPLTGSP 1135
DB 1076 ABSEAGSVDFPDGLGMGAKGLQSLPTHDPSPRLRYSDDPVLPLSEEDGVAPLTGSP 1135
OY 1136 QPEYVQNPVPRQPPSPREGRPLPAARPAATLERAKTISPGNGVYKDFAFGAVENE 1195
DB 1136 QPEYVQNPVPRQPPSPREGRPLPAARPAATLERAKTISPGNGVYKDFAFGAVENE 1195
OY 1196 YLTPGGAAPQHPPPAPSPAFNLYYMDQDPPERGAPSTFKGPTAENPEYGLDVPV 1255
DB 1196 YLTPGGAAPQHPPPAPSPAFNLYYMDQDPPERGAPSTFKGPTAENPEYGLDVPV 1255

```

RESULT 2  
 TERNU  
 protein-tyrosine kinase (BC 2.7.1.112) new precursor - rat  
 C:Species: Rattus norvegicus (Norway rat)  
 C:Date: 31-Dec-1988 #sequence\_revision 31-Dec-1988 #text\_change 11-Jun-1999  
 C:Accession: A24562; A61204  
 R:Barmann, C.I.; Hung, M.C.; Weinberg, R.A.  
 Nature 319, 226-230, 1986  
 A:Title: The new oncogene encodes an epidermal growth factor receptor-related protein  
 A:Reference number: A24562; MUID:86118662; PMID:945311  
 A:Accession: A24562  
 A:Molecule type: mRNA  
 A:Residues: 1-1260 <BAR>  
 A:Cross-references: EMBL:X03362; NID:956745; PIDN:CAA27059.1; PID:955746  
 R:Masu, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Con  
 R:Gingras 12, 1975-1978, 1991  
 A:Title: Direct DNA sequencing of the rat new oncogene transmembrane domain reveals:  
 2-thiazolyl]formamide or N-methyl-N-nitrosourea.  
 A:Reference number: A61204; MUID:92035293; PMID:1682063  
 A:Accession: A61204  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 637-663, 'V', 665-702 <MAS>  
 A>Note: authors translated the codon GCA for residue 25 as Val  
 C:Genetics:  
 A:Gene: neu  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos  
 F:1-19/Domain: signal sequence #status predicted <SIG>  
 F:20-1260/Product: protein-tyrosine kinase neu #status predicted <MAT>

8-680/Domain: transmembrane #status predicted <TM>  
 3-588/Domain: protein kinase homology <KIN>  
 1-739/Region: protein kinase ATP-binding motif  
 1,191,263,535,576,634/Binding site: carbohydrate (Asn) #status predicted  
 1/Binding site: phosphate (Thr) #status predicted  
 38/Active site: Lys #status predicted  
 32,1227,1253/Binding site: phosphate (Tyr) #status predicted

very Match 85.2%; Score 5805; DB 1; Length 1260;  
 est Local Similarity 85.3%; Pred. No. 9,4e-243; Indels 12; Gaps 5;  
 atches 1076; Conservative 58; Mismatches 116;

1 MELALCRWGLLALLPFGASTVCTGDMKRLPASPEHLDMRLHYGCCVYQGNL 60  
 4 MELAAWCRWGLLALLPFGASTVCTGDMKRLPASPEHLDMRLHYGCCVYQGNL 63  
 61 ELLYPTNASTSFLODIOEVQGVYLAHNVQVPLQRLIRYRGVQLPEEDNALAVLNG 120  
 64 ELLYPTNASTSFLODIOEVQGVYLAHNVQVPLQRLIRYRGVQLPEEDNALAVLNG 123  
 121 DPLNNTTPVT-GASPGGLRELQRLSLTEILKGGVLLQRPOLCYODTILMKDI FHKNNOL 179  
 124 DPQDNVASTPGRTPGRLRELQRLSLTEILKGGVLLQRPOLCYODTILMKDI FHKNNOL 183  
 180 ALTLIDNRSRACHPCSPKCKSGRCWGESSEDQSLRTVQAGGACRCKGPLETDCCHQ 239  
 184 APVIDITNRSRACHPCSPKCKSGRCWGESSEDQSLRTVQAGGACRCKGPLETDCCHQ 243  
 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTEFSPNPEGRYTFGASCVTAC 239  
 244 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTEFSPNPEGRYTFGASCVTAC 303  
 300 PYNLSTDVSGCTVCPPLHNOEVLAEDGTORCEKSKPCARVYGLQMGVYKANSKFI 359  
 304 PYNLSTDVSGCTVCPPLHNOEVLAEDGTORCEKSKPCARVYGLQMGVYKANSKFI 363  
 360 TELEFAGCKKIGSLAFLEPESFDGDPASNTAPLOEQLQVFTLEIIGVYISAMPDL 419  
 364 NVQEPFGCKKIGSLAFLEPESFDGDPASNTAPLOEQLQVFTLEIIGVYISAMPDL 423  
 420 PDLSTVFQNLQVIRGRILHANGSLTQIGISMLGRSLRELSGSLALHNNHLCFYHT 479  
 424 RDLSTVFQNLQVIRGRILHANGSLTQIGISMLGRSLRELSGSLALHNNHLCFYHT 483  
 480 VPMDOLEFRNHQALHTANPEDE-CVGEGLACHQACGHCMPGPTQCVNSQSLRQ 538  
 484 VPMDOLEFRNHQALHTANPEDE-CVGEGLACHQACGHCMPGPTQCVNSQSLRQ 543  
 539 ECVECEGVLOGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCAAHYDSSCVA 598  
 544 ECVECEGVLOGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCAAHYDSSCVA 603  
 599 RCPGCVKPDLSYMPWKPEDEGAQCPINCTHSCVDLDDKGPAPQASPLTISYSAV 658  
 604 RCPGCVKPDLSYMPWKPEDEGAQCPINCTHSCVDLDDKGPAPQASPLTISYSAV 663  
 659 VGILLVVLGVFGILLIKRQCKIRKTYMRRLLOETLEVEPLTSGAMPNOAKRIKET 718  
 664 BEVLLFLILVVLGVFGILLIKRQCKIRKTYMRRLLOETLEVEPLTSGAMPNOAKRIKET 723  
 719 ELRKVVLGSGAFGFENNFVSWLAVP-----KVASHLEVLRENTSPKANKEILDEAYV 773  
 724 ELRKVVLGSGAFGFENNFVSWLAVP-----KVASHLEVLRENTSPKANKEILDEAYV 778  
 774 MAGVSPVYSRLIGICTSTVQVLTQIMBYGCLLDHRENGRLSGODLLNMCQIAKEM 833  
 779 MAGVSPVYSRLIGICTSTVQVLTQIMBYGCLLDHRENGRLSGODLLNMCQIAKEM 838  
 834 SYLEDVLYVRDLAARNVLYKSPNHKAITDFGLARLIDETETNADGGKVPKMMALES 893  
 839 SYLEDVLYVRDLAARNVLYKSPNHKAITDFGLARLIDETETNADGGKVPKMMALES 898  
 894 ILRRFTHQSDVMSGYVWELMTFGAKPYDGI PAREI PDLLEKGERLPQPPICITIDVYM 953

DB 899 ILRRFTHQSDVMSGYVWELMTFGAKPYDGI PAREI PDLLEKGERLPQPPICITIDVYM 958  
 QY 954 INWKCMIMIDSECRPPRELIVSEFSNMARDPQRFVYIQNEDLGASPLDSTFYRLJLEDD 1013  
 DB 959 INWKCMIMIDSECRPPRELIVSEFSNMARDPQRFVYIQNEDLGASPLDSTFYRLJLEDD 1018  
 QY 1014 MGDVDAEYLVQCGFFCPDPAPGAGVNHRRHSSSTRSGGDLTLLEFSEEPAPS 1073  
 DB 1019 MGDVDAEYLVQCGFFCPDPAPGAGVNHRRHSSSTRSGGDLTLLEFSEEPAPS 1078  
 QY 1074 PLAPSEGASDVDDGLGMGAAGKGLSLPHTDPSPIQRSEDPVPLPSETDGVAPLTC 1133  
 DB 1079 PLAPSEGASDVDDGLGMGAAGKGLSLPHTDPSPIQRSEDPVPLPSETDGVAPLTC 1138  
 QY 1134 SPQPEYVNDVPRQPPSPREGPLPAPRAGATLERAKTSLPKQNGVYKDVAFGAVEN 1193  
 DB 1139 SPQPEYVNDVPRQPPSPREGPLPAPRAGATLERAKTSLPKQNGVYKDVAFGAVEN 1198  
 QY 1194 PEYLTPOGGAPOHPPPAPFAFDNLYYDQDPPERGAPESTKQPTTANPEYLGIDV 1253  
 DB 1199 PEYLTPOGGAPOHPPPAPFAFDNLYYDQDPPERGAPESTKQPTTANPEYLGIDV 1258  
 QY 1254 PV 1255  
 DB 1259 PV 1260

RESULT 3  
 148161  
 P-185 precursor - golden hamster  
 C:Species: Mesocricetus auratus (golden hamster)  
 C:Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #ext\_change 18-Jun-1999  
 C:Accession: 148161  
 R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishikawa, A.; Title: Cloning and activation of the Syrian hamster neu proto-oncogene.  
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.  
 A:Accession: 148161  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-1254 <RES>  
 A:Cross-references: GB:D16295; NID:G493236; PID:BAA03801.1; PID:G747595  
 C:Genetics:  
 A:Gene: neu  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP  
 F:718-983/Domain: protein kinase homology <KIN>  
 F:726-734/Region: protein kinase ATP-binding motif

Query Match 85.1%; Score 5795.5; DB 2; Length 1254;  
 est Local Similarity 84.9%; Pred. No. 2.4e-242; Indels 11; Gaps 4;  
 Matches 1070; Conservative 67; Mismatches 112;

1 MELALCRWGLLALLPFGASTVCTGDMKRLPASPEHLDMRLHYGCCVYQGNL 60  
 1 MELAAWCRWGLLALLPFGASTVCTGDMKRLPASPEHLDMRLHYGCCVYQGNL 63  
 61 ELLYPTNASTSFLODIOEVQGVYLAHNVQVPLQRLIRYRGVQLPEEDNALAVLNG 120  
 64 ELLYPTNASTSFLODIOEVQGVYLAHNVQVPLQRLIRYRGVQLPEEDNALAVLNG 123  
 121 DPLNNTTPVTGASPGGLRELQRLSLTEILKGGVLLQRPOLCYODTILMKDI FHKNNOL 180  
 124 DPQDNVASTPGRTPGRLRELQRLSLTEILKGGVLLQRPOLCYODTILMKDI FHKNNOL 183  
 180 ALTLIDNRSRACHPCSPKCKSGRCWGESSEDQSLRTVQAGGACRCKGPLETDCCHQ 239  
 184 APVIDITNRSRACHPCSPKCKSGRCWGESSEDQSLRTVQAGGACRCKGPLETDCCHQ 243  
 240 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTEFSPNPEGRYTFGASCVTAC 239  
 244 CAAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDTEFSPNPEGRYTFGASCVTAC 303  
 300 PYNLSTDVSGCTVCPPLHNOEVLAEDGTORCEKSKPCARVYGLQMGVYKANSKFI 359  
 304 PYNLSTDVSGCTVCPPLHNOEVLAEDGTORCEKSKPCARVYGLQMGVYKANSKFI 363  
 360 TELEFAGCKKIGSLAFLEPESFDGDPASNTAPLOEQLQVFTLEIIGVYISAMPDL 419  
 364 NVQEPFGCKKIGSLAFLEPESFDGDPASNTAPLOEQLQVFTLEIIGVYISAMPDL 423  
 420 PDLSTVFQNLQVIRGRILHANGSLTQIGISMLGRSLRELSGSLALHNNHLCFYHT 479  
 424 RDLSTVFQNLQVIRGRILHANGSLTQIGISMLGRSLRELSGSLALHNNHLCFYHT 483  
 480 VPMDOLEFRNHQALHTANPEDE-CVGEGLACHQACGHCMPGPTQCVNSQSLRQ 538  
 484 VPMDOLEFRNHQALHTANPEDE-CVGEGLACHQACGHCMPGPTQCVNSQSLRQ 543  
 539 ECVECEGVLOGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCAAHYDSSCVA 598  
 544 ECVECEGVLOGLPREYVNAHCLPCHPECPQNGSVTCFQPEADQCAAHYDSSCVA 603  
 599 RCPGCVKPDLSYMPWKPEDEGAQCPINCTHSCVDLDDKGPAPQASPLTISYSAV 658  
 604 RCPGCVKPDLSYMPWKPEDEGAQCPINCTHSCVDLDDKGPAPQASPLTISYSAV 663  
 659 VGILLVVLGVFGILLIKRQCKIRKTYMRRLLOETLEVEPLTSGAMPNOAKRIKET 718  
 664 BEVLLFLILVVLGVFGILLIKRQCKIRKTYMRRLLOETLEVEPLTSGAMPNOAKRIKET 723  
 719 ELRKVVLGSGAFGFENNFVSWLAVP-----KVASHLEVLRENTSPKANKEILDEAYV 773  
 724 ELRKVVLGSGAFGFENNFVSWLAVP-----KVASHLEVLRENTSPKANKEILDEAYV 778  
 774 MAGVSPVYSRLIGICTSTVQVLTQIMBYGCLLDHRENGRLSGODLLNMCQIAKEM 833  
 779 MAGVSPVYSRLIGICTSTVQVLTQIMBYGCLLDHRENGRLSGODLLNMCQIAKEM 838  
 834 SYLEDVLYVRDLAARNVLYKSPNHKAITDFGLARLIDETETNADGGKVPKMMALES 893  
 839 SYLEDVLYVRDLAARNVLYKSPNHKAITDFGLARLIDETETNADGGKVPKMMALES 898  
 894 ILRRFTHQSDVMSGYVWELMTFGAKPYDGI PAREI PDLLEKGERLPQPPICITIDVYM 953



ature 309, 270-273, 1994  
 Title: ATP-stimulated interaction between epidermal growth factor receptor and superoxide  
 Reference number: A38023; MUID:8419154; PMID:6325948  
 Contents: annotation; receptor activity  
 Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA  
 Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C.J.  
 Cell 59, 33-43, 1989  
 Title: Functional independence of the epidermal growth factor receptor from a domain  
 Reference number: A3331; MUID:9003233; PMID:2790960  
 Contents: annotation; internalization; signal  
 Comment: Binding of EGF to the receptor leads to internalization of the EGF-receptor  
 C:Genetics:  
 A:Gene: GDB:EGFR  
 A:Cross-references: GDB:120610; OMTN:131550  
 A:Map position: 7p12.3-7p12.1  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phospho  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:25-1210/Product: EGF receptor #status predicted <EXT>  
 F:25-645/Domain: extracellular domain repeat <EEL>  
 F:75-300/Domain: EGF receptor extracellular domain repeat <EE2>  
 F:390-600/Domain: transmembrane #status predicted <TM>  
 F:646-668/Domain: intracellular #status predicted <INT>  
 F:659-1210/Domain: protein kinase homology <KIN>  
 F:10-975/Domain: protein kinase homology  
 F:718-726/Region: protein kinase ATP-binding motif  
 F:999-1046/Region: coated-pit mediated internalization signal  
 F:1047-1210/Region: inhibitory  
 F:128, 175, 352, 413, 444, 528, 603/Binding site: carbohydrate (Aan) (covalent) #status predic  
 F:745/Active site: Lys #status experimental

Query March 44.5%; Score 3035; DB 1; Length 1210;  
 Best Local Similarity 48.7%; Pred. No. 1, se-123;  
 Matches 619; Conservative 176; Mismatches 359; Indels 116; Gaps 25;

11 LLLALLPPGAA-STVCTGTMTKLRLPASPETHLMDLRLXGCCVQVQGNLELYLPTN 68  
 14 LLLALCPASRLSEKKVCCGTSNLTQLGTFEDHFLSLQSMFENCEVLLGLLEITYQRN 73  
 69 ASLSFLDIDIVQVQVLIHNOVRQVPLRIRVGTOLFEEDNALVALDNGDPLNNTTP 128  
 74 YDLFLKTIQEVAGVLIANTVERIFLENQIRGMYYENSVALVLSYD----- 126  
 129 VTGASPGRLRELRLSTLEIKGVLIORNPOLCYQDTIMKDI FHKNNQALTLIDTNR 188  
 127 ---ANKTKELKPLRMNLEILHGAVRFENNPAICVESIQWRDIVSSDFISNMSMDFQNH 183  
 189 SRACHPCSPMCKGSRCKWSESEDCSLRTVCAGCA-RCKGPIPTDCHEQCAAGCTGP 247  
 184 LGSCKCKDPSCKNGSCWGEENCCQKTKITICAGCCSGCRCKSPSDCHNCAAGCTGP 243  
 248 KHSDDLACLHEHSGICELCPALVTYNTDFESMPNDEGRYTFGACVTAACPYNVSTD 307  
 244 RSDSLVCKKFKDEATCKTCTCPMLNNTTYQMDVNEPGKYSFATCCKKCPRYVVYTD 303  
 308 VESCTLVCPHNOVTAEDGTORCEKCSKPCARVCYGLCMQVTKNSKFIGITELE-PAG 366  
 304 HSCVACGADSYEM-EEDGVKCKCKCEPCRCVNCIGDIGEFK-DISISANTNIRKFN 361  
 367 CKKIFSLAFLESPFGDPASNTAPLOPEOLQVETLEETITGLYISAMPDLSPLSVQ 426  
 362 CTISISGLHLLPFAFGDSFTTPPLDPOBLDLKVKITIGLLILQAMPKRTDLHAP 421  
 427 NLQVIRGLIHNGAVSLTLOGIGISMLGRSELGSLALHNHNLCFVHTVPMDOJF 486  
 422 NLEIRGTQHOGFLAVVSLNITSIGRLSEISDGDVILSGNNKLCYANTINKKLF 481  
 487 RNPHQALITANPEDECVGEGLACHOLCARHGWGMPGPPQCNVCSQFLRGQECVECRV 546  
 482 GTSQCKTKIISNRNENCKATGVCHALCSPECGWGPFPDVCSCNVNSGRCDVCKTL 541  
 547 LOGPREYVNAHCLPCHPECPQNGSVTCFGEADQVACAHYKDPFPCVAPRSGVAP 606

542 LEGBPREFVENSECIQCHPECLPQAMNITCTGRGENDTQCAYHIDGPHCVKTCFAGVWG 601  
 607 DLSTYPMKFPPEBEGACQCPINCTHSCVYDLDKGCAPAEORASPLTISVAVG---ILL 663  
 602 ENNTL-VWKYADAGHYCHLCHNCTYGCCTGPELCEGPTNGKIP--SIATGVGALLLL 658  
 664 VVVLGVVFGILLKRRQCKIRKYMRLLOETELVEPLTPSGAMPQAKRIILKELELV 723  
 659 VVALGIG---LHRRRHIVRKTRRLLOERELVEPLTPSGAPNALRIILKELEFKKI 715  
 724 KVLGSAFGFNNFTVSFWLRVPKVSASHLEV---LRENTSPKANEILDEAVVAVGVS 779  
 716 KVLGSAFGF---TVYKGLWIPGEKVKIPVAIKELREATSPKANEILDEAVVAVGVS 771  
 780 PYVSLIGLCTSVOLVLTQMPYGLLDHRENRRLGSGDILLNMCQIAKGSYEDV 839  
 772 PHVRLIGLCTSVQILITQMLPFGCLLDVREHKONGSQYLNNMCVQIAKGNVLEDR 831  
 840 RLVRDLAARVNLVKSBNHYKITDFGLALLIDETVHADGGKVPKMALESILRRF 899  
 832 RLVRDLAARVNLVKTQHVKITDFGLAKLGAEEKHYHAGKVPKMALESILRRY 891  
 900 THQSDVMSYGVTVVEMLTFGAKKPYDGI PAEIPDLLEKGERLPQPICTIDVYMIWKW 959  
 892 THQSDVMSYGVTVVEMLTFGSKPYDGI PAEISISILEKGERLPQPICTIDVYMIWKW 951  
 960 MIDSCRRPREFLSEFSRMADEPQFVYIQ-NEDLGASPLDSTFTYHSLLEDMDGLY 1018  
 952 MIDASRRKRFELLIEFSGKMRDQRYLVIGDSERMLPSPDTSIFYALDEEDMDVV 1011  
 1019 DAEELVPOQGFCCPDPAPGAGKWHNRSSSTRSGGDLTLGEPSEEARSPPLAS 1078  
 1012 DAEELVPOQGF---SSPSTRPPLSS 1037  
 1079 EAGSDVPPDGDGMGAAGLQSLPTHPDSPLOQYSEDPVLPSET-DGYVAPLTCSEQ 1136  
 1038 LSAITSN-NSVACTIDRNLQSCPIKEDSFLOYSDFGALTEDSIDDTFL-----PV 1089  
 1137 PEYVQPDVPPQPPSPREGPLPARPAGATLERAKTSLSGKNGVYKDVAFGAVENPEY 1196  
 1090 PEYVQPDVPPQPPSPREGPLPARPAGATLERAKTSLSGKNGVYKDVAFGAVENPEY 1138  
 1197 L-TPGGAAAPQHPHPAPFADNLYWYQ-----DP-----PERGAPSPSTFGK 1239  
 1139 LNTVQ-----PFCVNSTFDSPAHMAQKSHQSLDNPDYQDQFFKPKAKRNCIFKG 1189  
 1240 TPTAENPEYL 1249  
 1190 S-TAENAEYL 1198

RESULT 5  
 A53183  
 epidermal growth factor receptor precursor - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999  
 C:Accession: A53183; A43818; S54942; A28941; S45325; I49643  
 R:Unetake, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.  
 Genes Dev. 8, 399-413, 1994  
 A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF recepto  
 A:Reference number: A53183; MUID:94170986; PMID:815255  
 A:Accession: A53183  
 A:Molecule type: mRNA  
 A:Residues: 1-1210 <LUE>  
 A:Cross-references: GB:U03425  
 R:Aviv, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.  
 Oncogene 6, 673-676, 1991  
 A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding  
 A:Reference number: A43818; MUID:91232866; PMID:2030916  
 A:Molecule type: mRNA  
 A:Residues: 1-714 <AVI>  
 A:Cross-references: GB:X59698

R.Eisinger, D.P.; Sertero, G.  
 submitted to the EMBL Data Library, June 1992  
 A:Reference number: S24942  
 A:Accession: S24942  
 A:Molecule type: mRNA  
 A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>  
 A:Cross-references: EMBL:Z12608  
 R:Heisermann, G.J.; Gili, G.N.  
 J. Biol. Chem. 263, 1152-1158, 1988  
 A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated  
 A:Reference number: A28941; MUID:88330814; PMID:3138233  
 A:Accession: A28941  
 A:Molecule type: protein  
 A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,  
 R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.  
 submitted to the EMBL Data Library, April 1994  
 A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Factor Receptor  
 A:Reference number: S45325  
 A:Accession: S45325  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-971, 'K', 973-1210 <VER>  
 A:Cross-references: EMBL:X76987; NID:948830; PIDN:CAA55587.1; PID:948831  
 R:Paria, B.C.; Das, S.K.; Andrews, G.K.; Dev, S.K.  
 Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993  
 A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse  
 A:Reference number: 149643; MUID:93126380; PMID:7678348  
 A:Accession: 149643  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 12-20, 22-132 <RES>  
 A:Cross-references: GB:L06864; NID:9193001; PIDN:AAA53029.1; PID:9567201  
 A:Accession: 149643  
 A:Gene: EGFR  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; growth factor receptor; kinase-related transforming protein; phosphop  
 F:1-24/Domain: signal sequence #status predicted <SIG>  
 F:648-670/Domain: transmembrane #status predicted <TM>  
 F:712-977/Domain: protein kinase homology <KIN>  
 F:720-728/Region: protein kinase ATP-binding motif  
 F:680, 695/Binding site: phosphate (Thr) (covalent) #status experimental  
 F:697, 1070, 1071/Binding site: phosphate (Ser) (covalent) #status experimental  
 F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental  
 F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental  
 F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental  
 Query Match 44.1%; Score 3006; DB 2; Length 1210;  
 Best Local Similarity 48.6%; Pred. No. 2,7e-122;  
 Matches 621; Conservative 169; Mismatches 367; Indels 120; Gaps 27;

364 -FAGCKKIFGSLAFLPESPFGDPASATAPLEOOLQVFTLEITCYLYISAMPDLPDL 422  
 358 HFKCTAISGDHLIPVAFKSDSFTTPPLDRELEILKTVKEITFLICAMPDWTDL 417  
 423 SVFQNLQVIRGLIHNGAVSLTQGLISWGLRLSRLSGSLALIHNTLCPVTVPM 482  
 418 HAFENLEIRGRKQHGQFSLAVVGINITSLGRSLKEISDDVLIISGRNLICANTIM 477  
 483 DQFRPHQALHTARPREDECVGBELACHQACAGHCMPPTOCVNSQFLRQCEVE 542  
 478 KKLFTGNQKTKIMNRAEKDCAVNHVNCPLSSBGCWGPBPDVSCQNSRGRCEVE 537  
 543 ECRVLQGLREYVNAHCLPCHPECOPOGVSATCFEPADQVACAHYKDPFCVYRCS 602  
 538 KNILGEEREFENSECIGCHPECLPQAMNITTCGRGDNICQCHYIDGHCVKTCGA 597  
 603 GYKPDLSYPIKFPPEBGAOCOPPLNCTHSCVYDLDDKCPRAEQRAPSLTISAVVGL 662  
 598 GIMGENNTL-VWKYADANNVCHLCHANCTYGCAGPGLQCEVWPSGPKIPSIATGVGL 656  
 663 LVVVLGVVFGI-LIKRQOKIRKYMRLLOETELVEPLTPSGAMPNOAMRLKETELR 721  
 657 LFTIV-VALGIGLFRKRRIIVKRTIRLLQRELVPLTPSGEAPNOAHLLKETERK 715  
 722 KYKVLGSGAFGNNTFVSWLBPVKVSAHLEV---LENTSPRANKIIDEAYVMAV 777  
 716 KIVLGSAGAFG---TVYKGLWIPGEKVKIPVAIKELSEATSPYANKIIDEAYVMAV 771  
 778 GSPYVRLIGLICTSTVQVLTQIMPYGCLLDHVRNRLGSGODLLNMQAKGSYLE 837  
 772 DNHVRRLIGLICTSTVQVLTQIMPYGCLLDHVRNRLGSGODLLNMQAKGSYLE 831  
 838 DVRLVHRLDAARNVLYKSPNHVKTDFGLARLLDIDETEHYADGKVPKXNALBSILR 897  
 832 DRLVHRLDAARNVLYKSPNHVKTDFGLARLLDIDETEHYADGKVPKXNALBSILR 891  
 898 RFTHQSDVWSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPPICITDVMIMVK 957  
 892 IYTHQSDVWSYGVTTWELMTFGAKPYDGIIPAREIPDLLEKGRLLPQPPICITDVMIMVK 951  
 958 CMNIDECPPRELVESESRMAPDPQREVVQV-NEDGAPAPLSTFPRSLLEDQMD 1016  
 952 CMNIDECPPRELVESESRMAPDPQREVVQV-NEDGAPAPLSTFPRSLLEDQMD 1011  
 1017 LVDAEEYLPQGGFFCPDPAPAGGVVHHRSSSTRSGGDLTLGLEPSEEARPSPLA 1076  
 1012 VVDADAEYLPQGGFF-----NSPT-----SRTPL 1037  
 1077 PSEGAGSDVFPDDDLGMAKAGLSLPTHDPSPLQYSEDPYPLPSET--DQVAPLITCS 1134  
 1038 SLSASNS---NSTVACINRNGSCRYKEDAFQRYSSDPTAVTEDNIDAFL----- 1087  
 1135 POBEVYNQDPVAPPPSPFREGPLPAPAPAGATLEBAKTLSPPKNSVVDVAFGAVENP 1194  
 1088 PVPEVYNQ-SVPRKAPAGSQNVHNPRLHP-----APRDLHYQN- PHSNAVGNP 1136  
 1195 EYL-TPQGAAPQHPPEAPFSPAPDNLTYWDO-----DP-----PERGAPSTF 1237  
 1137 EYLNIAQ-----PTCLSSGFSPALMIOKQSHQMSLDNDYQODFFPKEETPNQIF 1187  
 1238 KGTPTAENPEYGLDVP 1254  
 1188 KG-PTRENAEYLRVAP 1203

RESULT 6  
 TVCHLV  
 epidermal growth factor receptor precursor - chicken  
 N:contains: protein-tyrosine kinase (EC 2.7.1.112) erbB  
 C:species: Gallus gallus (chicken)  
 C:date: 28-Feb-1986 #sequence\_revision 05-May-1995 #text\_change 04-Feb-2000  
 C:accession: A27720; A00643

R.Lax, I.J. Johnson, A.J. Howk, R.J. Sap, J.J. Bellot, F.J. Winkler, M.J. Ulrich, A.J. Vennet  
 Mol. Cell. Biol. 8, 1970-1978, 1988  
 A>Title: Chicken epidermal growth factor (EGF) receptor: cDNA cloning, expression in mou  
 A:Reference number: A27720; PMID:88261272; PMID:3260329  
 A:Accession: A27720  
 A:Molecule type: mRNA  
 A:Residues: 1-1223 <LAX>  
 A:Cross-references: GB:M20386  
 A:Nilsen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rottman, F.M.; Crittenden, L.B.; Raines, M  
 Cell 41, 719-726, 1985  
 A>Title: C-erbB activation in ALV-induced erythroleukemia: novel RNA processing and pro  
 A:Reference number: A00643; PMID:85228222; PMID:2988784  
 A:Accession: A00643  
 A:Molecule type: mRNA  
 A:Residues: 585-1223 <NIL>  
 A:Cross-references: GB:M10066  
 C:Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: alternative splicing; ATP; autophosphorylation; glycoprotein; growth factor  
 C:Keywords: signal sequence #status predicted <SIG>  
 F:1-30/Domain: signal sequence #status predicted <SIG>  
 F:31-1223/Product: epidermal growth factor receptor #status predicted <NAT>  
 F:31-654/Domain: extracellular #status predicted <EXT>  
 F:31-307/Domain: EGF receptor extracellular domain repeat <EB1>  
 F:397-610/Domain: EGF receptor extracellular domain repeat <EB2>  
 F:655-677/Domain: transmembrane #status predicted <TM>  
 F:678-1223/Domain: intracellular #status predicted <INT>  
 F:719-984/Domain: protein kinase homology <KIN>  
 F:727-735/Region: protein kinase ATP-binding motif  
 F:136-202,280,361,370,422,575,580,615,635/binding site: carbohydrate (Thr) (covalent) #  
 F:192,650/binding site: carbohydrate (Ser) (covalent) #status predicted  
 F:687/binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted  
 F:754/Active site: Lys #status predicted  
 F:1100,1183,1208/binding site: phosphate (Tyr) (covalent) (by autophosphorylation) #stat

Query Match 43.8%; Score 2983.5; DB 1; Length 1223;  
 Best Local Similarity 47.5%; Pred. No. 2,66-121;  
 Matches 619; Conservative 174; Mismatches 354; Indels 155; Gaps 29;

QY 8 RWGLLALLPPGAA-----STQVCTGTDMKLRAPSPETHLMDLHLYOGGQVGNLE 61  
 DB 13 RGAALVLLLLGVALCSAVEEKVCGGTNNKLTQLGHVDHFTSLQRMVNCSEVLSNLE 72  
 QY 62 LTLPTNALSIFQDIOEQVGYLLAHNQVQVPLQRLKLVKSTQLEFEDNVALYLNGD 121  
 DB 73 IYVEHNRDLTFLKTLQEVAGVYLLANVADVLPENLDIIRGNVLYDMSFALAVLSNH 132  
 QY 122 PLNNTPTVGTASPGGLRELQSLSTELKGVLIQRNPOLCYODTILMKOIFHKNQAL 181  
 DB 133 -NNKTO-----GRELPMKRLSEILNGVYKISNPKLCMMDTVLMNDIIDTSRK-PL 182  
 QY 182 TLID-TNRSDACHPCSPCKGRSGWESSEDCQSLTRVCAAGCA-RCKGRLPTDCHEQ 239  
 DB 183 TVLDFASNLSSCKCKHPNCTEDHCWAGQNCQTLTKVICAQCSGRCKGVPDCHNQ 242  
 QY 240 CAAGCTGPHSDCLACLFHNSGICELHCPALVTYVTDPEEMPNPEGRTGASCTYAC 299  
 DB 243 CAGGCTGPRSDCLAKRKRFRDQATCKOTPRVLYNPPTTYQMDVPEGRYSFGATVCEC 302  
 QY 300 PNYVISTDVSGCTLVCPHNOEYVADGTQRCCKSPCARCYGLGMOYIRANSKFTGI 359  
 DB 303 PHNYVYTDHSGCVRSQNTDTYEV-EENGVRCKCKCGCLSKVNGIGELGILLS-INA 360  
 QY 360 TELE-PAGCKKIFGSLAFIPESFDGPASNTAPLQEQVETLEITGYIYISAMPDS 418  
 DB 361 TINDSRKCTKINGDVSLPVAFLGAPFTKPLDPKLDVRRVKEISGFLLIQAMPDN 420  
 QY 419 LPDLASYFQNLQVIRGLIHNGAYSLTQGLIGISWGLRSLRSLGSLALIHNTLQFVH 478  
 DB 421 AFDVLAFFENLEIRGTQKHQGYSLAVVNLKIQSLGRLSKISDGLAIMKRLCTAD 480  
 QY 479 TVPWQDLFRNPQALHTANRPDECEVGGGLACHQICARGHCWGGPRTQCVNCSQFLRQG 538

DB 481 TMAWRSPLATOSQKTKIIONENKNDCTDRHVCDPLGSDVCGMGPGPHCSRCFFRSKX 540  
 QY 539 ECVESCRVLQGLPRFYVYVARNRCLPCHPCCQONG----SVTPGPRADQVCANRKPDPF 595  
 DB 541 ECVKCNLLQGEPRFERDSKCLPCHSCLQVNSTAYVTTQSGPRDCKMCAHIDDPH 600  
 QY 596 CVARCPSPKPDLSMPYMKFPDEBGAQCPPICTHSCVLDLDDGCAEORASPLTSIV 655  
 DB 601 CYKACPAVLGENDTL-VKXADANAVQLCHPCTROCKPGLGECG---NGSTPBLA 656  
 QY 656 SAVV-GILLVVLGVFGILIKRQOKIRKYMRLLOETELVEPLTPSGAMPQAKRI 714  
 DB 657 AGVVGGLLCLVVGIGLIGLYLRR-HYRKRTLRLLQERELVEPLTPSGAPPAQAHRI 715  
 QY 715 LKETEELRVKYLGSAGAFGNFTYSFWLRVAKVASHEV---LRENTSKANKELLDE 770  
 DB 716 LKETEELRVKYLGSAGAFG---TVKGLWIEGKRVKIPVAKELRENTSKANKELLDE 771  
 QY 771 AYVAGVSPYVSRLLGICLTSTVQLVTLQMPYGLDHDVRENRGLSGODLWCMQIA 830  
 DB 772 AYVAGVSPYVSRLLGICLTSTVQLVTLQMPYGLDHDVRENRGLSGODLWCMQIA 831  
 QY 831 KGMSTLEVRVLRDLARNVLYVSPNHVKTITDGLARLDIDETEHADGKVPYKMA 890  
 DB 832 KGMVLEERRLVHRDLARNVLYVTPQHVKITDGLAKLGADEKEYAEGKVPYKMA 891  
 QY 891 LESIIRRFTHQSDVSGVTVWELMTFGAKPYGIPAREIPDLLEKGERLPORPCTID 950  
 DB 892 LESIIRRYTHQSDVSGVTVWELMTFGAKPYGIPAREIPDLLEKGERLPORPCTID 951  
 QY 951 VYMIWVKCMTIDSECRPRELVEFSFARMDFORFVYIQ-NEDLGAPSPDSTFFYRLL 1009  
 DB 952 VYMIWVKCMTIDSECRPRELVEFSFARMDFORFVYIQ-NEDLGAPSPDSTFFYRLL 1011  
 QY 1010 EDDMGDLVDAEYLYVPOGGFCECDPARGACGMVHRRSSSTSGGDLTLGLPSEEE 1069  
 DB 1012 EEDMEDIVDAEYLYVPOGGFCECDPARGACGMVHRRSSSTSGGDLTLGLPSEEE 1038  
 QY 1070 APRSPL-----APSEGADVDFDGLGKAAGLQSLPTHPSPLOYSEDPVLPSET 1124  
 DB 1039 -SRPILSSLSATENNSATNID-----RNQGHVYRDSFVQKSSPTGNFLES 1089  
 QY 1125 -DGVVAPLTSPOPEYVNPQVRPQPSREGLPAPAPAGATLERAKTLPKNGVYK 1182  
 DB 1090 IDDGFL-----PAPEVYNQ--LMPKXPS-----TAMQONQLYN 1120  
 QY 1183 DVF-----AFGAVNEEYLLPQGAAPQHPHAPFAPFNLYVMD-- 1225  
 DB 1121 NISLTAISKLPMSDRYQNSHTAVDNEPYL-----NTNQSPLAKTVESSPWYIQSG 1172  
 QY 1226 -----DPE-----RGAPSTFKGTPTAENEYLGLDVP 1254  
 DB 1173 NHQNLNDPVOYQDFLPNETKPNGLKVPAAENEYLYVYAP 1214

RESULT 7  
 A47253  
 A>Title: Human epidermal growth factor receptor, HER4 - human  
 C:Date: 22-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 18-Jun-1999  
 C:Accession: A47253  
 R:Plowman, G.D.; Culouscou, J.M.; Whitney, G.S.; Green, J.M.; Carlon, G.W.; Foy, L  
 Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993  
 A>Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epide  
 A:Reference number: A47253; PMID:93189574; PMID:8383326  
 A:Accession: A47253  
 A>Status: preliminary; not compared with conceptual translation  
 A:Molecule type: nucleic acid  
 A:Residues: 1-1308 <FLO>  
 A:Cross-references: GB:U07868; NID:G337359; PIDN:AA59446.1; PID:G337360  
 A>Note: sequence extracted from NCBI backbone (NCBI:126842)  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology



C:Keywords: ATP; growth factor receptor  
F:716-981/Domain: protein kinase homology <KIN>  
F:724-732/Region: protein kinase ATP-binding motif

Query Match 41.9%; Score 2852.5; DB 2; Length 1308;  
Best Local Similarity 44.1%; Pred. No. 1,2e-115;  
Matches 596; Conservative 187; Mismatches 385; Indels 183; Gaps 31;

```

Qy 9 WGLLALLPPGAA-----STQCTGTMKRLRLPASPEHMLDMKHLVQGCQVQVGNLELY 64
Db 8 WWWSLVLAAGTVQPSDSQVACATEKSLSDLEQYALRKRYENCCEVVGNELEITS 67
Qy 65 LPTNASLFLQDIQEVGYLIANQVQVPLQRLIRVSTQLEFEDNYALAVLNDQDPAN 124
Db 68 IEHNRDLSPRSAREVGYLVALLNQFRYPLEMLRIRGTXYEDRYALALFLNKRKG 127
Qy 125 NTPFVTGASPGGLRELRLSLTELKGGVLIQRNPQLCYODTLKMDIFHKNNQALLTL 184
Db 128 NF-----GLOELGLKRLTELINGVYVDQNKFLCYADTIHMQDIYRNWPMSNLTLV 178
Qy 185 DYNRRACHPCSPKCKGRGCGESSEDCQSLRTVCAGGC-ARCKGPLETDCCHQCCAG 243
Db 179 STNGSSGCGRCHKSCG-RCMGPTENHCQTLRTVCAGQCDGRCYGYVSDCCHRCAG 227
Qy 244 CTGPRHSDCLACHFRHSGICELHCPALVTYNDTPRESMPEGRTFGASCYTACPRY 303
Db 238 CSGPKDPCACAMFNDSGACVTCQCPQTFVYNPTFQLEHNFNAKYTGAFCYCKKPHNF 297
Qy 304 LSTVGSCTLVCPHNOEYTAEDGTQRCCKSKPCARVCGLMQVYKANKRFTIGTELE 363
Db 298 V-VDSSCVRACPSKMEV-ENNGIKMKCKCTDLCRACGIGTGLSMAGTYDSSNIDK 355
Qy 364 FAGCKKIFGSLAPRESFDDPASNAPLOPEQLQVETLEITGYLYISAWPDSLDLS 423
Db 356 FINCTKINGNLIFLVTGIDHPYALAIPEKLNVRVREITGLINIGSMPPNMTDS 415
Qy 424 VFQQLVIRGRILHNGAVSLTLQGLISWLGSRSLRSLSGALLHHNHTLCPVHVPMD 493
Db 416 VFSVLVTGGVLSJSLILKQGLTSLQFQSLKISGNIYITDMSLCTYHITNMT 475
Qy 484 QLEFNPHQALLHTANREDECVSGGLACHQJCARGHCWGPPTQCVNCSQFLRGCEVE 543
Db 476 TLFSTINGRIIVIRNRAKENCETAEGVGNHLCSDGCGWGPQCLSCRRFSGR.CIES 535
Qy 544 CRVQLGPREVVARHLLPCHPEQF-QNGSVTCFGEALQCAAHKRPPECVACRS 602
Db 536 CNLYDGEFREHENSICVECDPQCKMEDGLTCHGFPNCTKCSHFKGPNCVCKCPD 595
Qy 603 GVKPDLSEYMPIMKFPDEGACQPCPINCTSHCVLDLDKGC-----PAEORASPL 651
Db 596 GLQGANSP--IFKXADPRECHPCPNCTQCGNGPTSHDCIYYPWTHGSHLPQHAR-TPL 652
Qy 652 TSIVSAYV-GILLVVYGVVFGILLIKRQOKIRKRYTKRLLQSTELVEPLTSGAMPNO 710
Db 653 --IAAGVIGGLFVLIVGLTFAYVVRKRSIK-KKRALRFL-STELEVEPLTPGTAPNO 708
Qy 711 QMRILKTELKRVKVLGSAFGFNNFVSEFLRVP-----KVASHLEVIRENTSPKAK 765
Db 709 QLRILKTELKRVKVLGSAFG-----IVYKGINWBEETIKIYVA-IKILNETTGKANY 763
Qy 766 EILDEAVYKAGVSPYVSRLLGICLTSTVOLVTLQMPYGLLDHVRNCRGLSQDLNWK 825
Db 764 EFMDEALIMASMDHPLVRLGLVCLSPITQVLMHGHGILLVEYHKKNISQILLNWK 823
Qy 826 CMOJAKMSYLEDEVLVHRDLAARNVLYKSPNHVKTDFGLARLDIDEDVYADGKYP 885
Db 824 CVOJAKMSYLEERLVRHDLAARNVLYKSPNHVKTDFGLARLDIDEDVYADGKYP 883
Qy 886 IKMWALESLRRRTFTHSGDVSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLP 945
Db 884 IKMWALESLIRKRTFTHSGDVSYGVTVWELMTGAKPYDGIIPAREIPDLLEKGERLP 943
Qy 946 ICTIDVYMWKCMWIDSECRPRFRELVSFSRMAARDPQGFVVIQND-LGPASPLDSTF 1004

```

```

Db 944 ICTIDVYMWKCMWIDSECRPRFRELVSFSRMAARDPQGFVVIQNDRLMLPSNDKSF 1003
Qy 1005 YRSLLEDDMDGLVDAAEELVYPQCGFCPPDPAPGAGVTHHRHSSSTRSGGDTLLGLE 1064
Db 1004 FQNLLEDDLEDDMDAAEELVYP-QAFNIPPP-----LYTSRRARIDSNRS-----EIGHS 1051
Qy 1065 PSEEEAFPS-----PLAP-SEGAGSDVFDGLGMGAA 1095
Db 1052 PPAPATPMSGNOFVTRDGFAAEGQVSPYPAPISTIPEAVAGATAEITDDSCNGTL 1111
Qy 1096 KGLQSLPTHDPSPPLQRYSEDPVPLPS-----ETDGYVAPLTCSPOPEYVQDPVBPQ 1148
Db 1112 RKPVAIPHQEDSGSTRYAGADPTVFAPERSPRGELDEEGYMTPMRDKPKQEYLNPEV 1167
Qy 1149 PPSPRGELPAPRPGATLLEAKTSLPQKGVVMDVFAFGCAVENPEVLTQGGAAAPQH 1208
Db 1168 -----ENPVSRR-----KNGDLQ-----ALDNPEYHNASNG----- 1194
Qy 1209 PPPA-----FSPAFDNLVYWDODPPERGA-- 1232
Db 1195 PPKADEYVNEPELYINTFANTLGAKEYLKNILSWPERAKKAFQDPDMHNSLPPRSTLQ 1254
Qy 1233 PSTFKGTPT-----AENPEYL 1249
Db 1255 HPDYLOEYSTKYFYKQNGRIRIYVANEPEYL 1285

RESULT 8
S06142
protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish
N:Alternate names: epidermal growth factor receptor homology; kinase-related transfo
C:Species: Xiphophorus maculatus (southern platyfish)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 18-Feb-2000
C:Accession: S06142; S13809
R:Witbrodt, J.; Adam, D.; Malitschek, B.; Meuweler, W.; Raulf, F.; Telling, A.; Ro
Nature 341, 415-421, 1989
A:Residues: 1-1166 <WIT>
A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing T
A:Reference number: S06142; MUID:90015140; PMID:2797166
A:Accession: S06142
A:Molecule type: DNA
A:Residues: 1-1166 <WIT>
A:Cross-references: EMBL:X16891; NID:g65290; PIDN:CAA34770.1; PID:g65291
R:Adam, D.; Meuweler, W.; Schattl, M.
Oncogene 6, 73-80, 1991
A:Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xipho
A:Reference number: S13807; MUID:91125882; PMID:1846957
A:Accession: S13809
A:Status: preliminary; translation not shown
A:Molecule type: DNA
A:Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>
A:Cross-references: EMBL:X55319; NID:g65284; PIDN:CAA39763.1; PID:g65285
C:Genetics:
A:Gene: mrk
A:Map position: Y
A:Antigens: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein;
F:1-23/Domain: signal sequence #status predicted <SIG>
F:16-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 37.6%; Score 2559; DB 1; Length 1166;
Best Local Similarity 43.9%; Pred. No. 4,6e-103; Indels 156; Gaps 34;
Matches 560; Conservative 168; Mismatches 392; Indels 156; Gaps 34;

Qy 4 AALCWGILLALLPPGAAS-----QVCTGTMKRLRLPASPEHMLDMKHLVQGCQVQVGN 59
Db 8 AALLQ--LLLVLSISRCSTDPDRKVCCTGNQMT---LDNHYLKMKNYSGGVLEN 62
Qy 60 LEITLTPNASLSFLQDIQEVGYLIANQVQVPLQRLIRVSTQLEFEDNYALAVLND 119

```



```

Db      63 LETTYQENODLSFLOSIOEVGYLIANNVSTIPVLNRLIRGNLYEGNFTLIWNS 122
Qy      120 GDELNTTPTVGASPGELRELOLRSLTEILKGVLIQIORNPOLCYODTILMKDIFHANNOL 179
Db      123 YCK-NPSSP--DYVQGLKQLOLSNLTILSGVAKSHNPLCNVETIMWMDIVKTSNP 179
Qy      180 ALTLIDTNSRAHPCSPMCKSGRCWGESSEDCQSLTRIVCAGGC-ARCKGPLPTDCHE 238
Db      180 TMLIPHAERQCKDHGCVNGSCWAPGPHQCKFTKLLCAEQCNRRGRGKPIDCCNE 239
Qy      239 QCAAGCTGPKHSDCLACLFHNHSGICELCPALVTNTDTEFSMPNPEGRTYFGASCVA 298
Db      240 HCGAGCTGPRATDCLACRDPNDGTCKDCTCPKXIYDIHQVNDPNKITYFGACVGE 299
Qy      299 CPYNYLSTVGSCTLVCPHNDVETADGTQRCCKSKPCQARVCGLGKQVIYKANSKFTG 358
Db      300 CPENYVYTE-GACVRCSCAGKMLEVD-ENKGRCKPCDGVCPKVCIDIGISL-SNTIAVN 356
Qy      359 ITFL-EPAGCKIFGSLAFLPESFGDDPASNTAPLQPELOVFEETEEITGYLYISAWPD 417
Db      357 STIRSFNSCTKINGDIIINRNSFEEDPHYKIGTMDPEHMLNLTITXELTGYLVIMWPE 416
Qy      418 SLPDLSEFONLOVIRGRILHNGAYS-LTLQGLISWLGSLRSLGSLALIHNTLTCF 476
Db      417 NMTSLSVFQNLTEIRGRTFSRGFSFVAVQVRLQWLGRLKEVAGNVILKNTLQRLY 476
Qy      477 VHTVPMQLFERNHQLLTANRPEDECEVGLAQHQLCARHGCWPGTQCVNGSQIFR 536
Db      477 ANINMRRLFRSDQIEYDART-----ENQTCNNESESDGCW-PGPTVCWSCILHYDR 528
Qy      537 GQCEVECEVLOGLPREYVNAHRLCPHPECOPONGSVCFGEADQCAVACHYADPPRC 596
Db      529 GRCVASCNLLQGEPEADVDGRCVQCHQELVQDLSLTCTYGRPNCKSKAHFQDGPQC 588
Qy      597 VAPCSGVKPLDSYPMIKFPDEEGACQPCPINCCHSCVLDLDDKGPABORASPLTISVS 656
Db      589 IPRCPHIGLGDGDTL-IMKYADKMGCCQPCQONCTQCGSGPGLSGRGD-IVSHSLAVG 646
Qy      657 AVVGLLVVVLGVFGILKBRQKIRKXTMRLLQETLVEPLTPSGMAMPQAKRIK 716
Db      647 LVSGLLITVIALIVLRLRRRIK-RKSTIRCLQEKELVPLTPSGAPQAPRLRIK 705
Qy      717 ETELRKYVLGSGAGFNNFTV-SFW-----LRVKSASHLEVRNTPKANKELL 768
Db      706 ETEFKDRVLGSGAGF---TYKGLMNDGENIRLPAV-----IKTLRAATSPKYNQDEL 757
Qy      769 DEAYWAGVSPVNSLLGICLTSTYQVLTQMLPFGCLLDHRENRGRIGSODLIANWQ 828
Db      758 DEAYWASVDHPHVCRLGICLTSAQVLTQMLPYGCLLDYRQHERICQGMWLNWCQ 817
Qy      829 IAKGMSYLEDVLRHDLAARVLYKSPHVKITPFGARLADI DETEYHADGKVPYIKM 888
Db      818 IAKGMNYLEERHLYHRDLAARVLLKNPHVKITDPGLSKLLTABEKVQADGKVPYIKM 877
Qy      889 MALESTLRRFTHQSCVMSYVTVWELMTFGAKPYDGPAREIPDLLEKERELEPPIC 948
Db      878 MALESTLQWTYTHQSCVMSYVTVWELMTFGSKPYDGPAREIPDLLEKERELEPPIC 937
Qy      949 IIVYIMVCKMIDSCPRPRELVSEFSRMAADQRFVYIQNEDLGASPLDSTFYNSL 1008
Db      938 IEVYMIILCKMIDBSRPRFRELVEFSQMAKDSRYLVIG---NLPSLSDRLFLRSL 994
Qy      1009 LEDDDMGDLVDAEYLVPQGFCCPDPAAGAGVYHRRSSSTRSGGDLTLGLPESEE 1068
Db      995 LSSDD--DYVDADEYLLPYKRI-----NKGGS----- 1019
Qy      1069 EAPRSPLASBEGAGSVFPGDGLMGAKAGLQSLPHDPSPLQRYEDPTV-PLBSETGY 1127
Db      1020 ---EPCLEPTGH-----PVRENSITLRNISTDPLQNALERDLGH 1055
Qy      1128 VAPLTCSPQPEYVNOPVNPP-----PSPRE-----GPLP-AAPAGATLTERAKTIS 1174
Db      1056 -----EYVNOGSETTSRLSDIYNPNVEDLTDGNGFVSSSQEAFETNFSRPEYLN 1105

```

```

Qy      1175 PKKNGVNDVEAFGAVENPEYLTPOGGAAPQHPHPAFAFNLVYWDODPBERGAP 1234
Db      1106 TNQNSL---PLVSGSMDDPDY---QAG-----YQAAF-----LPQTGALT 1140
Qy      1235 STFKCTPFAENPEYLG 1250
Db      1141 GNGMFLPAENLEYLG 1156

RESULT 9
A36223
Kinase-related transforming protein (erbB3) (EC 2.7.1.-) precursor - human
C/Dates: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C/Accession: A36223; 159164
R/Kruse, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.
Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A/Title: Isolation and characterization of ERB3, a third member of the ERBB/epider
A/Reference number: A36223; MUID:90083234; PMID:2687875
A/Accession: A36223
A/Status: Preliminary
A/Molecule type: mRNA
A/Residues: 1-1342 <XPA>
A/Cross-references: GB:M29366
R/Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A/Title: Molecular cloning and expression of another epidermal growth factor recept
A/Reference number: 159164; MUID:90311312; PMID:2164210
A/Accession: 159164
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-559; 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A/Cross-references: GB:M34309; NID:G183990; PIDR:AAA35979.1; PID:G306841
C/Genetics:
A/Gene: GDB:ERBB3; HER3
A/Cross-references: GDB:119880; OMIM:190151
A/Map position: 12q13-12q13
C/Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h
C/Keywords: ATP; phosphotransferase
F/707-972/Domain: protein kinase homology <KIN>
F/715-723/Region: protein kinase ATP-binding motif

Query Match      33.9%; Score 2307.5; DB 2; Length 1342;
Best local similarity 39.4%; Pred. No. 3.5e-92;
Matches 519; Conservative 196; Mismatches 460; Indels 143; Gaps 36;

Qy      10 GLIALALPPGAA--STQVCTGTDMKRLRPASPETHMLBHLVYCGVQGNLELYLPT 67
Db      11 GLTFLARSGSEVNSGAVCGTGLNGLSVTGDAENQYQTLVLYRCEVWGNLELYLGH 70
Qy      68 NASLSFIQDIOEVGYLIANNVROVPLORLRIYRGTQLFEDNYALAVALDNGDPLNNT 127
Db      71 NADSLFQWIREVGYLVANNRSTPLPLRLRYRGTQYVDGFAIFVM-----LNYNT 125
Qy      128 PVTGASPGARELOLRSLTEILKGVLIQIORNPOLCYODTILMKDIFHANNQLALTLDTN 187
Db      126 ---NSSHALRQLRLTLEILSGVYIEKDKLCHNDITDWRITVDRD---AEIVKQ 178
Qy      188 RSRACHPCSPMCKSGRCWGESSEDCQSLTRIVCAGGC-ARCKGPLPTDCHEQCAAGCTG 246
Db      179 NGRSCPCHEVCKG-RCMGPGSEDCQTLTITICAPQCNHGCFGNPNQCHDEAGGCSG 237
Qy      247 PKHSDCLALCFHNHSGICELCPALVTYNTDTEFSMPNPEGRTYFGASCYTACRYNYLST 306
Db      238 PDDTCFACRHFNDGSCVPRCPQLVYNLTQLLENPHTKYQYGGVCAASCPHNFV-V 296
Qy      307 DVGSCTLVCPHNDVETADGTQRCCKSKPCARVCGLGKQVIYKANSKF--IGITFL- 363
Db      297 DQSCVACPPDKMEVD-KNGLKMCBCGGLCPKACGCTG-----SSRQGYDSSNIDG 350
Qy      364 FAGCKIFGSLAFLPESFGDDPASNTAPLQPELOVFEETEEITGYLYISAWPDLSDLS 423

```

Db 351 FVACTKILGNLDELITGLNGDPWPKH:PALDPEKLNFRVREITGYLNQSMPPHMFNS 410  
 Qy 424 VFQNLQVIRGRILHNGAYS-LTLOG:GISWLS:RSURELSSGAL:HNHNLHCFYTVPM 482  
 Db 411 VPSNLTTIGRSILNRGFSLLIMKNLVISLGRSLKEISAGIYISANRQLCYHSLWM 470  
 Qy 483 DQLEFRPHQALLHTA-NRPEDECVGEGLAGHOLCARHGCWGPPTOCVNCQPLRGQECV 541  
 Db 471 TKVLRGPTERLDIKHNRRPRDCVAEGKVCPLSSGCGPFGQCLSCRNYSRGVCV 530  
 Qy 542 EECRVLOGLPREYVNAHCLPCHPEOCOPNGSVTCGPPEADQCAVACHYKDPFCVACRP 601  
 Db 531 THCNFLNGEPREERHAEHCEFSCHPEOCMEGTATCGSSGSDTAQCAHFRDGHCVSSCP 590  
 Qy 602 SGVCPDLSTYMPIMKFPDEEGACOPCPINCTHSCVDDDDKCPAEGRA---SPITISVA 657  
 Db 591 HGVLG--AKGPITYKYPVQNECRPHCHENCQSGKPELODQGLQTLVLGKTHLTALIV 648  
 Qy 658 VGVLLVVLGVVFGILIKRQOKIR-KYTMRLLOETELVEPLTSGAMPNOQNRILK 716  
 Db 649 IAG--LVVIFMMLGGTFELVWRGRRIQNKRAMRYLERGESIEPLDPS-EKANKVLARIEK 705  
 Qy 717 ETELKRVKVLGSGAFNNFTVSMFLRVPKVSASH---LEVIRENTSPKANKELIDEAY 772  
 Db 706 ETELKRVKVLGSGVFG---TVKGVWIPGESIKIPVCIKYEDSGKSGFOAVTDHML 761  
 Qy 773 VMAGVSPVYSRLIGLITSTVOLVTLQMPYGLLDHVRNRRGLSODLLWMCQIAGK 832  
 Db 762 AIGSLDHAHIVRLICPGSSQLQVLPGLSLDHLVHRGALPQLLLMVGVIAGK 821  
 Qy 833 MSYLEDVRLVHRPLAARNVLKSPSNHYKIDPGLALLDIDETEHADGKVPKIMALE 892  
 Db 822 MYLLEHGMWHRMLAARNVLKSPSOVAVADFGVADLPDDQLLYSEKAPITKMALE 881  
 Qy 893 SILRRFTHOSDVMSYGVTVWELMTFGAKPYDGIIPAREIPDLLEKEGRLPOPICTIDV 952  
 Db 882 SIHFQKVTQSDVMSYGVTVWELMTFGAEFYAGRLAEVVDLLEKEGRLAOPICIDV 941  
 Qy 953 MIVKCMMDISECRPRELVSFESMARDPORFVITQNDLCPA---SPLDSTFYSLL 1009  
 Db 942 MWVKKCMMDENIRPEKELANFTMARDPRIYLVIKRS--GPIAGPGBEPLTNKXL 1000  
 Qy 1010 EDDMDGLVDAEYLVQCGFPCPDPAAGAGWVHHRSSSTRSGGDLTGLDEP-SEE 1068  
 Db 1001 EYVELEFELDLDLLEBED---NLATTIGSALSLPVGTLNRP 1041  
 Qy 1069 EABRSPLABEGAGSDVFGDGLMGAAGLQSLPTH-DSPLORYSEDPVPLP 1121  
 Db 1042 RGSQSLSPSSGY-MPMNQNLGESCOCESAVSGSSERCPVSLH-----PMRGCLAS 1094  
 Qy 1122 SETDGYVA-----PLTCSPOPE---YNNQPDVRPOPPSPREGP 1156  
 Db 1095 ESSEGHVTSGEALQEVSMCRSRSSRSPRPGDAYSQHSQSLTPTPLSPPELEEE 1154  
 Qy 1157 -----LPAARPAGATLERAKTLP-GKNGV-----KDVFAFGAVENPEYLPQGA 1203  
 Db 1155 DVNGVMPDTHLKGTPSSREGTLLSVGLSLCTEBEDD-----EYEVNMRNRH 1206  
 Qy 1204 AQPGRPPAPSPAFNDLYWD-----QDPRKGAAPTSTKGPITANPEYL 1249  
 Db 1207 SP-PHPRPSSLLEELGYMDVGSLSLSTGTSQCPHLHPVIMPTAGTTPDEDEYEM 1263

## RESULT 10

JC4387

epidermal growth factor receptor homolog precursor - rat

N:Alternate names: ErbB3 protein; HER3 protein

C:Species: Rattus norvegicus (Norway rat)

C&gt;Date: 17-Jan-1996 #sequence, revision 19-Apr-1996 #text, change 13-Nov-1998

C/Accession: JC4387

R/Hellyer, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.

Gene 165, 279-284, 1995

A&gt;Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protein.

A:Reference number: JC4387; MUID:96096535; PMID:8522190  
 A:Accession: JC4387  
 A:Molecule type: mRNA  
 A:Residues: 1-1339 <HEL>  
 A:Cross-references: GB:U29339; NID:9915389; PID:9915390  
 A:Experimental source: Liver  
 A>Note: The authors translated the codon AAC for residue 369 as Thr and GTT for res  
 C:Comment: This protein is a functional heregulin receptor that transduces signals  
 A:Gene: ErbB3  
 C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase h  
 C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane prote  
 F:1-19/DNA: signal sequence #status predicted <SIG>  
 F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>  
 F:640-659/DNA: transmembrane #status predicted <TMM>  
 F:705-970/DNA: protein kinase homology <XIN>  
 F:713-721/Region: protein kinase ATP-binding motif  
 F:939,1051,1156,1196,1219,1257,1259,1273,1286,1325/Binding site: phosphate (7y

## Query Match

Best Local Similarity 39.5%; Pred. No. 1.8e-88;  
 Matches 509; Conservative 173; Mismatches 440; Indels 167; Gaps 37;

Qy 3 LAALCRWGLLALPFGCA---STQVCTGDMKRLRPASEPETHLDMRLHYOGQOVQGN 59  
 Db 7 LQVLC---FLSLARGSEMGNSQAVCPGTINGSLVGDADNOYQTLTKLYEKEVWGN 62  
 Qy 60 LETLTPTNASLFLDIOEVOGYLLAHNOVROVPLORLIRIVGTOLFEDYVNAVLAVDN 119  
 Db 63 LEIVTGNADLSFLQWIREVTAYLVVAMNESVLPRLNLVAVGTQYDQKFAIFVM-- 120  
 Qy 120 GDPLNTPTVTGASFGGLRELQSLTEILKGVLIQNPOLCYQDTLLMKDIFHKNNQL 179  
 Db 121 ---LNVNT---NSSHALRQLKFTQLEITELSGVYIEKNDLCHMDITDMRDIYVR-- 170  
 Qy 180 ALTLIDTRSSAACHPCSMCKSGKCGWSSSDCSLRTVTAGGC-ARCKELPTDQHE 238  
 Db 171 GAELIVKKNAGNCPCEHEVCCKG-RCKWGGPDDCQILTKTICAPQNGCFQPNPNOCCHD 229  
 Qy 239 QCAAGCTGPKHSDCLACHFNHSGICELHCAALTYNTDTEESMNPREGRTFGASCYTA 298  
 Db 230 ECAGSGSPDQDTCACGRFNDGACVRCGEPLVYNKLTQLLEPNHTKQYGVCVAS 289  
 Qy 239 CPYNYLSTDVSGTILVCPILHNOEYTAEDGTORCEKSKPCARVYGL--GMQYIKANSKF 356  
 Db 230 CPNHFV--VDQTFVCYACAPCPDMEVD-KIGLKPCPCGGLCPKACGEGSGSGRYQTVSSN 347  
 Qy 357 IGITLFPAGCKKTFGSLAFPLPESFDGDPASNTAPLOEOLQVEFTEITGYLYISAWP 416  
 Db 348 ID---GFVNCITKILGNLDELITGLNDPWHKIPALDPEKLNFRVREITGYLNQSM 403  
 Qy 417 DSLPDLVFNQLQVIRGRILHNGAYS-LTLOG:GISWLS:RSURELSSGAL:HNHNLHCF 475  
 Db 404 PHMHFVSFNSLTTIGRSLYNGFSLIMKNLVISLGRSLKEISAGIYISANQDLC 463  
 Qy 476 FVHTVPMQQLFRNPHQALLHTA-NRPEDECVGEGLAGHOLCARHGCWGPPTOCVNCQ 534  
 Db 464 YHSLNMTRLLRGSEERLDIKYRPGECCLABEKVCDPLSSGCGPFGQCLSCNRY 523  
 Qy 535 LRQGEVCECRVLOGLPREYVNAHCLPCHPEOCOPNGSVTCGPPEADQCAVACHYKDP 594  
 Db 524 SREGVCTHGNFLGEPREHAEHCEFSCHPEOCMEGTATCGSSGSDTAQCAHFRDGH 583  
 Qy 595 FCVACRSGVAPDLSYMPIMKFPDEEGACOPCPINCTHSC--VLDLKGCPAEGRASPLT 652  
 Db 584 HCVNSCPGILG--AKGPITYKYPDAQNECRCHENCQOGCGPELODQGLQCAEVLMSKPH 641  
 Qy 653 SIVAVVGLLVVLGVVFGILIKRQOKIR-KYTMRLLOETELVEPLTSGAMPNOQAO 711  
 Db 642 LVIAVVG--LAVILMLIGSFLVWRGRRIQNKRAMRYLERGESIEPLDPS-EKANKVL 698  
 Qy 712 WRLKEFELKRVKVLGSGAFNNFTVSMFLRVPKVSASH---LEVIRENTSPKANKEL 767

Db 699 ARIFKTELEKLVGSGVGF---TVHKGIWPEGSIKIPVICIKYIEDKSRGSEQAV 754  
 Qy 768 LDEAYVAVAGSVYSSRLIGICTSTVQVLTQIMPYGCLLDHYRENRGRGLSGDILNMCM 827  
 Db 755 TDHMLAVGSLDHAIVRLIGICPGSSIQVLTQVLTPLGSLLDHYKORRETLPGQLLNMGV 814  
 Qy 828 QIAKGMVYLEDVNLVHNDLAARVLYKSPNHVKITPGLARLDDIETETHADGKVPK 887  
 Db 815 QIAKGMVYLEDVNLVHNDLAARVLYKSPNHVKITPGLARLDDIETETHADGKVPK 874  
 Qy 888 WMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPIC 947  
 Db 875 WMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIPDLLEKGERLPQPPIC 934  
 Qy 948 TIDVYVIMVKCMWIDSECRPRRELVSERPMARDPQRFVIVIONEDLGASPLDSTFYRS 1007  
 Db 935 TIDVYVIMVKCMWIDSECRPRRELVSERPMARDPQRFVIVIONEDLGASPLDSTFYRS 991  
 Qy 1008 LLEDDMDGLVDAEEYLVPOQGFPCPDPAAGAGVWVHRRSSSTRSGGDLTLGLPBE 1067  
 Db 992 VLTTEL---GEAELEPEL-----DLDDLEABE 1017  
 Qy 1068 E-----EAPRSPPLASEG-----AGSDVFDGDLGMAKGLQ 1099  
 Db 1018 EGLATSLGSLSLPTGTLTRPGSGSLSPSSGYMNMQSLGEACLDGAVLGGREGFSR 1077  
 Qy 1100 SLTTPDSPLOQRSEDPVLPSETDGV---APL-----TC-----SPOPE---Y 1139  
 Db 1078 PISLH-PIDRGR-----PASESSGHTVGSFAELIOEKVSVCRSSRSRSPFRGDSAY 1129  
 Qy 1140 VNQPDVPRPPSPREGE-----LPAARPAATLERAKTLSP-GKNGV----- 1181  
 Db 1130 HSGRHSLLTFTVTLSPGLPEEDGNGVMPDTHLRGASSRECTLLSVGLSSVLTGREED 1189  
 Qy 1182 KDVFAFGAVENVEYLTPOGGAAPQHP 1210  
 Db 1190 ED-----EEYEMNRKRGRSP-PRP 1209

## RESULT 11

TVEVIV

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian leukosis virus  
 N:Contains: amino end of gag protein; env protein fragment; protein-tyrosine kinase  
 C:Species: avian leukosis virus, ALV  
 C>Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 11-Jun-1999  
 C:Accession: B00643; A00643  
 R:Hilgen, T.W.; Maroney, P.A.; Goodwin, R.G.; Rotman, F.M.; Ciltenden, L.B.; Raines, M.  
 Cell 41, 719-726, 1985  
 A>Title: C-erbB activation in ALV-induced erythroblastosis: novel RNA processing and pro  
 A:Reference number: A00643; MUID:85228222; PMID:2988784  
 A:Accession: B00643  
 A:Molecule type: mRNA  
 A:Residues: 1-698 <NLS>  
 A:Cross-references: GB:M10066; GB:M13881; NID:g211749; PID:AAA48763.1; PID:g211750  
 A:Note: in Genbank entry CKEB8BF, release 109.0, the source is designated as Gallus gal  
 C:Comment: This protein is synthesized as a gag-env-erbB protein.  
 C:Genetics:  
 A:Gene: gag-env-erbB  
 A:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specific P  
 F.1-6/Product: gag protein (fragment) #status predicted <GAG>  
 F.7-59/Product: env protein (fragment) #status predicted <ENV>  
 F.60-698/Product: protein-tyrosine kinase #status predicted <ERB>  
 F.194-459/Domain: protein kinase homology <KIN>  
 F.202-210/Region: protein kinase ATP-binding motif  
 F.229/Active site: Lys #status predicted

## Query Match

Best Local Similarity 24.3%; Score 1655.5; DB 1; Length 698;  
 Matches 361; Conservative 81; Mismatches 145; Indels 133; Gaps 20;

578 GPEADQCVACAHYKDPFCVACPSGVKFDLSYMBIMKPPDEAGACOPICHTSCYDLD 637

Db 60 GP--DHCKCAHFIDGPHCVKACPAVLGENDTL-VKKYADANAVCOLCHPNCTRGCKGP 116  
 Qy 638 DDGCPAPCARASPLTSIYSAVY-GILLVVLGVGVGILIKRQOKIRKYTMRLDQETEL 696  
 Db 117 GLEGGP---NSGKTSIAAGVGGILCLVYVGLGLVLRNR-HIVRKTLRRLIQEDEL 172  
 Qy 697 VEPLTPSGAMNQAOIRILKETELRKVKVLSGAFGNFTVSPFLRVKVSASHLEY-- 754  
 Db 173 VEPLTPSGAPQAOHLRIKETEFKVKVLSGAFG---TVYKGLWPEGSIKIPVAI 228  
 Qy 755 --LRNTSPKANKELDEAYVAVAGSVYSSRLIGICTSTVQVLTQIMPYGCLLDHYRE 812  
 Db 229 KELRATSPKANKELDEAYVAVAGSVYSSRLIGICTSTVQVLTQIMPYGCLLDHYRE 288  
 Qy 813 NRGRLGSDLLNMCMQIAKGMVYLEDVNLVHNDLAARVLYKSPNHVKITPGLARLDDI 872  
 Db 289 HNDNIGSQVLLNMCMQIAKGMVYLEDVNLVHNDLAARVLYKSPNHVKITPGLARLDDI 348  
 Qy 873 DETETHADGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGAKPYDGIPIAREIP 932  
 Db 349 DEKEYHAGGKVPKIMMALESILRRFTHQSDVMSYGVTVWELMTFGSKPYDGIPIAREIS 408  
 Qy 933 DLLEKGERLPQPPICITIDVYVIMVKCMWIDSECRPRRELVSERPMARDPQRFVIVION 991  
 Db 409 SVLEKGERLPQPPICITIDVYVIMVKCMWIDSECRPRRELVSERPMARDPQRFVIVION 468  
 Qy 992 EDLGASPLDSTFYRSLEDDMDGLVDAEEYLVPOQGFPCPDPAAGAGVWVHRRSSS 1051  
 Db 469 ERMHLPSPDTSKFYVTLMEEDMEDIVDAEYLVHQEFF-----NSPS 512  
 Qy 1052 TRSGGDDLTLGLEPSEERAPRSP-----ASEGASVFPFGDLGMAKGLQSLPTIDP 1106  
 Db 513 T-----SRPLSSISATSNNSATNCID-----RNGGHPVED 546  
 Qy 1107 SPLQRYSEDPVLPSET---DGYVAFLTCSPOPEYVNPQPDVPRPPSPREGLPARAPAG 1164  
 Db 547 SFVQYSSDPDPTGNFLEESIDGFL-----PAPRYNQ--LMPKPS----- 585  
 Qy 1165 ATLERAKTLSPGKNGVQVDF-----AFGAVENVEYLTPOGGAAPQHP 1209  
 Db 586 -----TAVVQNCIYNNISLTAISKLPMSRYQNSHSTAVDNEYL-----NTNQ 629  
 Qy 1210 PPAPSPAPDNLYYMDQ-----DPPE-----RGAPPTFGTTPAENVEYLGIDVP 1254  
 Db 630 SPLAKTVFESSPPYMIQSGNHQINIDNPYQODFLPNETKPGLLKVPAAENVEYLVRAAP 689

## RESULT 12

TYVUH

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain 1  
 C:Species: avian erythroblastosis virus 18-Apr-1984 #text\_change 11-Jun-1999  
 C>Date: 18-Apr-1984 #sequence\_revision 18-Apr-1984 #text\_change 11-Jun-1999  
 C:Accession: A00644; A38022  
 R:Yamamoto, T.; Nishida, T.; Miyajima, N.; Kawai, S.; Ooi, T.; Toyoshima, K.  
 Cell 35, 71-78, 1983  
 A>Title: The erbB gene of avian erythroblastosis virus is a member of the src gene f  
 A:Reference number: A00644; MUID:84028539; PMID:6313229  
 A:Accession: A00644  
 A:Molecule type: DNA  
 A:Residues: 1-604 <YAM>  
 A:Cross-references: GB:X01216; NID:g209676; PID:AAA42400.1; PID:g209678  
 R:Debutre, B.; Henry, C.; Benatissa, M.; Biserte, G.; Claverie, J.M.; Saule, S.; Mart  
 Science 224, 1456-1459, 1984  
 A>Title: Sequencing the erbB gene of avian erythroblastosis virus reveals a new type  
 A:Reference number: A38022; MUID:84223957; PMID:6328658  
 A:Accession: A38022  
 A:Molecule type: DNA  
 A:Residues: 1-28, 'W', '30-139', 'F', '141-145', 'V', '147-152 <DEB>  
 A:Cross-references: GB:X02006  
 C:Genetics:  
 A:Gene: erbB  
 A:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C:Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-specif

F:130-395/Domain: protein kinase homology <KIN>  
 F:138-146/Region: protein kinase ATP-binding motif  
 F:165/Active site: Lys #status predicted

## Query Match

23.4%; Score 1592; DB 1; Length 604;  
 Best Local Similarity 50.0%; Pred. No. 1e-61;  
 Matches 347; Conservative 77; Mismatches 136; Indels 134; Gaps 18;

```

QY 587 CAHYKDPFCVAPCPGVPDLSTYMPIMKPFDEGACOPCPINCTHSCVLDLDCGCPAEQ 646
DB 3 CAHFDGHCYKACAPAGVLENDTL-VKRYADANAQOLCHPCTCKCGKRGEGC--- 58
QY 647 RASPLTSIVSAV-V-GILLVVLGVFGILIRKROQIRKRYMRRLQETELVEPLPSGA 705
DB 59 NGSKTSPSIAGVVGGLCLVVGGLGIGLYLR-R-HIVKRTLRRLQERELVEPLPSGE 117
QY 706 MPNOAMPILKETELRKVYLGSGAFGNNFTVSFMLRVPKVSASHLEV---LRENTSP 761
DB 118 APNOHRLILKETELRKVYLGSGAFG---TLYKGLMIFEGEKVKIPVAIKELREKATSP 173
QY 762 KANKEILDEAVYVAGSPVYSRLIGICTSTVQLTQLMFYGCLLDHVENRGLSQD 821
DB 174 KANKEILDEAVYVAGSPVYSRLIGICTSTVQLTQLMFYGCLLDHVENRGLSQD 223
QY 822 LLMCMQIAKMSYLEVDRLVHRDLAARNVLYKSPNNVKTIDREGLRLDIDETRYADG 881
DB 234 LLNMCVQIAKMSYLEVDRLVHRDLAARNVLYKSPNNVKTIDREGLRLDIDETRYADG 293
QY 882 GKVPILKMALESILRRFTTHQSVDVSYGVTVLMTFGAKPYDGIIPAREIPDLLEKGERL 941
DB 294 GKVPILKMALESILRRFTTHQSVDVSYGVTVLMTFGAKPYDGIIPAREIPDLLEKGERL 353
QY 942 POPICITIDVYIMVCKMIDSECRPRELVSFSKMDPQCFVYIQ-N-NEDIGPASP 1000
DB 354 POPICITIDVYIMVCKMIDSECRPRELVSFSKMDPQCFVYIQ-N-NEDIGPASP 413
QY 1001 DSTFYRSLLEDMDGDLVDAEELVLPQGFPCPDPAAGAGVYHRRSSSTRGCGDLT 1060
DB 414 DSKFYRSLLEDMDGDLVDAEELVLPQGFPCPDPAAGAGVYHRRSSSTRGCGDLT 449
QY 1061 LGLEPSEEARSPFL---APSEGAGDVDFGDLGMAKGLSLPTHPSPLORYSED 1115
DB 450 LGLEPSEEARSPFL---APSEGAGDVDFGDLGMAKGLSLPTHPSPLORYSED 491
QY 1116 PTVPLPSET-DGVAPLTCSPQPEYVQPDVPRPQSPREGPLPAARPAATLERAKTL 1173
DB 492 PTVPLPSET-DGVAPLTCSPQPEYVQPDVPRPQSPREGPLPAARPAATLERAKTL 524
QY 1174 SPGRNGVVKDVFARF-----GAVENPEYLTPOGGAAPQHPPPAFS 1214
DB 525 SPGRNGVVKDVFARF-----GAVENPEYLTPOGGAAPQHPPPAFS 1214
QY 1215 PAFDNLVYWDPPERKAPSTFKGTFAENPEY 1248
DB 571 TVFESSPYWISQGNH-----INLDNPDY 594

```

## RESULT 13

353745  
 protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus  
 C:Species: avian erythroblastosis virus  
 C>Date: 03-Mar-1994 #sequence revision 26-May-1995 #text\_change 28-Feb-1997  
 C/Accession: S35745  
 R:Vennstrom, B.  
 submitted to the EMBL Data Library, March 1993  
 A:Accession: S35743  
 A:Molecule type: DNA  
 A:Residues: 1-544 <VEN>  
 A:Cross-references: EMBL:X12707  
 C/Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: ATP; oncogene; phosphotransferase; transforming protein; tyrosine-speci  
 F:135-400/Domain: protein kinase homology <KIN>  
 F:143-151/Region: protein kinase ATP-binding motif  
 F:170/Active site: Lys #status predicted

## Query Match

22.6%; Score 1541; DB 2; Length 544;  
 Best Local Similarity 52.7%; Pred. No. 1.4e-59;  
 Matches 333; Conservative 71; Mismatches 128; Indels 100; Gaps 17;

```

QY 578 GREADCCVCAHYKDPFCVAPCPGVPDLSTYMPIMKPFDEGACOPCPINCTHSCVLD 637
DB 1 GP-DHCKCAHFDGHCYKACAPAGVLENDTL-VKRYADANAQOLCHPCTCKCGK 57
QY 638 DKGCPAEQASPLTSIVSAV-V-GILLVVLGVFGILIRKROQIRKRYMRRLQETEL 696
DB 58 GLEGCP---NGSKTSPSIAGVVGGLCLVVGGLGIGLYLR-R-HIVKRTLRRLQERL 113
QY 697 VEPLTPSGAMPQAOQRIKETELRKVYLGSGAFGNNFTVSFMLRVPKVSASHLEV--- 754
DB 114 VEPLTPSGAMPQAOQRIKETELRKVYLGSGAFG---TLYKGLMIFEGEKVTIPVAI 169
QY 755 -LRENTSPKANKEILDEAVYVAGSPVYSRLIGICTSTVQLTQLMFYGCLLDHVE 812
DB 170 KETRENTSPKANKEILDEAVYVAGSPVYSRLIGICTSTVQLTQLMFYGCLLDHVE 229
QY 813 NGRIGSODLLNMCQIAKMSYLEVDRLVHRDLAARNVLYKSPNNVKTIDREGLRLDI 872
DB 230 HKDNIGSODLLNMCQIAKMSYLEVDRLVHRDLAARNVLYKSPNNVKTIDREGLRLDI 289
QY 873 DETRYADGCKPIKMALESILRRFTTHQSVDVSYGVTVLMTFGAKPYDGIIPAREIP 932
DB 290 DETRYADGCKPIKMALESILRRFTTHQSVDVSYGVTVLMTFGAKPYDGIIPAREIP 349
QY 933 DLLEKGERLPOPICITIDVYIMVCKMIDSECRPRELVSFSKMDPQCFVYIQ-N-991
DB 350 SVLEKGERLPOPICITIDVYIMVCKMIDSECRPRELVSFSKMDPQCFVYIQ-N-409
QY 992 EDLGRASPDLSTFYRSLLEDMDGDLVDAEELVLPQGFPCPDPAAGAGVYHRRSS 1051
DB 410 EDLGRASPDLSTFYRSLLEDMDGDLVDAEELVLPQGFPCPDPAAGAGVYHRRSS 453
QY 1052 TRSGGDLTLGLEPSEEARSPFL---APSEGAGDVDFGDLGMAKGLSLPTHP 1106
DB 454 TRSGGDLTLGLEPSEEARSPFL---APSEGAGDVDFGDLGMAKGLSLPTHP 481
QY 1107 SPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYVQPDVPRPQSPREGPLPAARPA 1166
DB 482 SPLQRYSEDPTVPLPSETDGVAPLTCSPQPEYVQPDVPRPQSPREGPLPAARPA 519
QY 1167 -LERAKTSPGRNGVVKDVFARF-----GAVENPEYLTPOGGAAPQHPPPAFS 1197
DB 520 AISKLPIDSRYN-----SHSTAVDNEEYL 544

```

## RESULT 14

500727  
 kinase-related transforming protein (erbB) (EC 2.7.1.-) - avian erythroblastosis vi  
 C:Species: avian erythroblastosis virus  
 C>Date: 01-Dec-1989 #sequence revision 01-Dec-1989 #text\_change 28-Feb-1997  
 C/Accession: S00727  
 R:Scotting, P.; Vennstrom, B.; Jansen, M.; Graf, T.; Beug, H.; Hayman, M.J.  
 Oncogene Res. 1, 265-278, 1987  
 A>Title: Common site of mutation in the erbB gene of avian erythroblastosis virus m  
 A:Reference number: S00727; MUID:88217326; PMID:2897102  
 A:Accession: S00727  
 A:Molecule type: DNA  
 A:Residues: 1-545 <SCO>  
 A:Cross-references: EMBL:X06943  
 C/Genetics:  
 A:Gene: erbB  
 C:Superfamily: epidermal growth factor receptor; protein kinase homology  
 C/Keywords: ATP; phosphotransferase  
 F:135-400/Domain: protein kinase homology <KIN>

F:143-151/Region: protein kinase ATP-binding motif

Query Match 22.5%; Score 1534; DB 2; Length 545;  
 Best Local Similarity 52.7%; Pred. No. 2.8e-59;  
 Matches 333; Conservative 70; Mismatches 129; Indels 100; Gaps 17;

QY 578 GPADQCACAHYKDPFCYACPCSGYKEDLSYMFYKPPDEGACQPCPCINCHSCVDL 637  
 DB 1 GP--DHCMKCAHFIDGPHCVKACPCAGVLGENDTL-VWKYADANAVCOLCHPNCRCGCKGSP 57  
 QY 638 DDGCPAEQASPLTSTVSAYV--GILLVVLGVVFGILLKRRQCKIRKYMRLLOETEL 696  
 DB 58 GLECCP--NGSTPSTIAAGVVGGLCLVVGIGIGLYLRR-HYVKRLRLRLQRELVLPSPG 113  
 QY 697 VEPULTSGAMPNOAKRLKTELRYKVLGSGAFGNFTVSPFWLVPKVSASHLEV-- 754  
 DB 114 VEPULTSGAMPNOAKRLKTELRYKVLGSGAFGNFTVSPFWLVPKVSASHLEV-- 169  
 QY 755 --LRENTSPKANEILDEAYVAGVSPYVSLIGTCTSTVQLTQMLPYGCLLDHRE 812  
 DB 170 KELREATSPKANEILDEAYVAGVSPYVSLIGTCTSTVQLTQMLPYGCLLDHRE 229  
 QY 813 NRGRLGSDLLNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKTIDFGIARLDI 872  
 DB 230 HKDNISQVILNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKTIDFGIARLDI 289  
 QY 873 DETEYHADGSKVPIKMALESILRRFTHOSDVSYGVTVWEIMTFGAKPYDGIAPREIP 932  
 DB 290 DEKEYHAEGGKVKIKMALESILRRFTHOSDVSYGVTVWEIMTFGAKPYDGIAPREIP 349  
 QY 933 DLEKGERLPPICTIDVYIMVWKCMIDSECRPRELVSFESRARDPQRFVYIQ-N 991  
 DB 350 SVLEKGERLPPICTIDVYIMVWKCMIDSECRPRELVSFESRARDPQRFVYIQ-N 409  
 QY 992 EDLGPASPLDSTFYRSLLBEDDMGDLVDAEYLVPOQGFCCPDPAFGAGMVHRRSSS 1051  
 DB 410 ERNHLPSFTDSKRYRTIMEBEDMEDIVDAEYLVPOQGF-----NSPS 453  
 QY 1052 TRSGGCDLTLGLEPSEEARSP-----APSEGAGSDVFDGDLGMAKGLQSLPHTDP 1106  
 DB 454 T-----SRTPLLSLSATSNNSATNCIDRNG-----H-- 481  
 QY 1107 SPLQRYSEDPVPLPSETDGYVAPLTCSPQPEYVNOQDVAPQPPSPREGPLPAPAPAGAT 1166  
 DB 482 -----PVAREDFL-----PAPRYVNO--LMPKRPSTAMVQNOIYVYISLT 519  
 QY 1167 -LERAKTLSPGKGVKGVKVPFAGGAVENPEYL 1197  
 DB 520 AISKLPMDSRYN-----SHSTAVDNPEYL 544

## RESULT 15

B44776

protein-tyrosine kinase (EC 2.7.1.112) erbB - avian erythroblastosis virus (strain ES4)

C/Species: avian erythroblastosis virus

C/Date: 28-Apr-1993 #sequence\_revision 28-Apr-1993 #text\_change 04-Feb-2000

C/Accession: B44776

R:Bruskin, A.; Jackson, J.; Bishop, J.M.; McCauley, D.J.; Schatzman, R.C.

Oncogene 5, 15-24, 1990

A/Title: Six amino acids from the retroviral gene gag greatly enhance the transforming P

A/Reference number: A44776; M0ID:9020603; PMID:1969616

A/Accession: B44776

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-540 &lt;BRU&gt;

A/Cross-references: GB:X52211

C/Genetics:

A:Gene: erbB

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C/Keywords: ATP; phosphotransferase; tyrosine-specific protein kinase

F:130-395/Domain: protein kinase homology &lt;KIN&gt;

F:138-146/Region: protein kinase ATP-binding motif

Query Match 22.3%; Score 1517; DB 2; Length 540;  
 Best Local Similarity 52.6%; Pred. No. 1.5e-58;  
 Matches 328; Conservative 70; Mismatches 127; Indels 98; Gaps 16;

QY 567 CAHYDPPFCYACPCSGYKEDLSYMFYKPPDEGACQPCPCINCHSCVDLDDKCPAEQ 646  
 DB 3 CAHFIDGPHCVKACPCAGVLGENDTL-VWKYADANAVCOLCHPNCRCGCKGSP 58  
 QY 647 RASPLTSTVSAYV--GILLVVLGVVFGILLKRRQCKIRKYMRLLOETELVEPLTPSGA 705  
 DB 59 NGSTPSTIAAGVVGGLCLVVGIGIGIGLYLRR-HYVKRLRLRLQRELVLPSPG 117  
 QY 706 MPNOAKRLKTELRYKVLGSGAFGNFTVSPFWLVPKVSASHLEV--LRENTSP 761  
 DB 118 APNOAKRLKTELRYKVLGSGAFGNFTVSPFWLVPKVSASHLEV--LRENTSP 173  
 QY 762 KANEILDEAYVAGVSPYVSLIGTCTSTVQLTQMLPYGCLLDHRENRGLSGD 821  
 DB 174 KANEILDEAYVAGVSPYVSLIGTCTSTVQLTQMLPYGCLLDHRENRGLSGD 233  
 QY 822 LLNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKTIDFGIARLDI DETEYHAD 881  
 DB 234 LLNMCQIAKMSYLEDVRLVHRDLAARNVLYKSPNHVKTIDFGIARLDI DETEYHAD 293  
 QY 882 GKVPKMALESILRRFTHOSDVSYGVTVWEIMTFGAKPYDGIAPREIPDLLEKGERL 941  
 DB 234 GKVPKMALESILRRFTHOSDVSYGVTVWEIMTFGAKPYDGIAPREIPDLLEKGERL 353  
 QY 942 POPPICTIDVYIMVWKCMIDSECRPRELVSFESRARDPQRFVYIQ-NEDLGPASPL 1000  
 DB 354 POPPICTIDVYIMVWKCMIDSECRPRELVSFESRARDPQRFVYIQ-NEDLGPASPL 413  
 QY 1001 DSTFYRSLLBEDDMGDLVDAEYLVPOQGFCCPDPAFGAGMVHRRSSS TRSGGCDL 1060  
 DB 414 DSKFYRTIMEBEDMEDIVDAEYLVPOQGF-----NSPST----- 449  
 QY 1061 LGLEPSEEARSP-----APSEGAGSDVFDGDLGMAKGLQSLPHTDPSPLQRYSED 1115  
 DB 450 -----SRTPLLSLSATSNNSATNCIDRNG-----H-- 476  
 QY 1116 PTVPLPSETDGYVAPLTCSPQPEYVNOQDVAPQPPSPREGPLPAPAPAGAT -LERAKTL 1174  
 DB 477 -----PVAREDFL-----PAPRYVNO--LMPKRPSTAMVQNOIYVYISLT 523  
 QY 1175 PGKGVKGVKGVKVPFAGGAVENPEYL 1197  
 DB 524 RYON-----SHSTAVDNPEYL 539

Search completed: July 22, 2003, 05:28:12  
 Job time : 30.2855 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 08:47:04 ; Search time 10.642 seconds

(without alignments)  
4891.279 Million cell updates/sec

Title: SEQ4-710-730-14

Perfect score: 6813  
Sequence: 1 MELALCRWGLLLALLPGCA.....TFKTPTAENPEYIGLDVAV 1255

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database: SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	6602	96.9	1255	ERB2_HUMAN	P04626 homo sapien
2	5811	85.3	1257	ERB2_RAT	P06494 rattus norv
3	5795.5	85.1	1254	ERB2_MESAU	P06053 mesocricetu
4	3033	44.5	1210	EGFR_HUMAN	P00533 homo sapien
5	3007	44.1	1210	EGFR_MOUSE	Q01279 mus musculu
6	2852.5	41.9	1308	ERB4_HUMAN	Q15303 homo sapien
7	2833	41.6	1308	ERB4_RAT	Q62856 rattus norv
8	2575.5	37.8	1167	XRK_XIPMA	P13388 xiphophorus
9	2244.5	34.0	1342	ERB3_HUMAN	P13860 homo sapien
10	2244.5	32.9	1339	ERB3_RAT	Q62799 rattus norv
11	1836	26.9	1426	EGFR_DROME	P04412 drosophila
12	1638.5	24.0	634	ERBB_ALV	P00534 avian leuko
13	1595	23.4	703	EGFR_CHICK	P13387 gallus gall
14	1592	23.4	604	ERBB_AVIER	P00535 avian eryth
15	1524	22.4	540	ERBB_AVIER	P11273 avian eryth
16	1217	17.9	1323	LT23_CAEL	P24348 caenorhabdi
17	1142.5	16.8	245	ERB2_MOUSE	P70424 mus musculu
18	693	10.2	1363	ILPR_BRALA	O02466 brachiole
19	656.5	9.6	1372	ILPR_MOUSE	P15208 mus musculu
20	656	9.6	1300	IRK_MOUSE	Q9W114 mus musculu
21	655	9.6	1382	INSR_HUMAN	P06612 homo sapien
22	653	9.6	1383	INSR_RAT	P15127 rattus norv
23	647	9.5	1297	IRR_HUMAN	P14616 homo sapien
24	644.5	9.5	1300	IRR_CAVPO	P14617 cavia porc
25	632	9.3	1477	HTK7_AYDAT	Q25197 hydra atten
26	630	9.2	1607	MITP_LYNST	Q25410 lynnaea stea
27	602	8.8	1367	IGIR_HUMAN	P08069 homo sapien
28	584	8.6	1373	IGIR_MOUSE	P240751 mus musculu
29	580.5	8.5	1370	IGIR_RAT	P06052 rattus norv
30	580	8.5	1114	RET_HUMAN	P07949 homo sapien
31	579	8.5	1141	RET_MOUSE	P54760 homo sapien
32	575	8.4	2146	INSR_DROME	P09208 drosophila
33	569	8.4	1390	INSR_AEDAE	Q93105 aedes aegypt

34	559	8.2	987	1	EPB4_MOUSE	P54761 mus musculu
35	551.5	8.1	984	1	EPB1_CHICK	Q07494 gallus gall
36	548.5	8.1	984	1	EPB1_RAT	P09759 rattus norv
37	544	8.0	902	1	EPB3_XENLA	Q91736 xenopus lae
38	542.5	8.0	984	1	EPB1_HUMAN	P54762 homo sapien
39	536.5	7.9	985	1	EPB4_XENLA	Q91571 xenopus lae
40	534.5	7.8	977	1	EPB2_MOUSE	Q03145 mus musculu
41	529.5	7.8	976	1	EPB2_HUMAN	P29317 homo sapien
42	525.5	7.7	985	1	EPB4_XENLA	Q91736 xenopus lae
43	525.5	7.7	1053	1	PAK1_CHICK	Q00944 gallus gall
44	524.5	7.7	986	1	EPB4_XENLA	Q91845 xenopus lae
45	523	7.7	1068	1	PAK1_XENLA	Q91738 xenopus lae

## ALIGNMENTS

```

RESULT 1
ID ERB2_HUMAN STANDARD, PRT, 1255 AA.
AC P04626;
DT 13-AUG-1987 (Rel. 05, Created)
DR 13-AUG-1987 (Rel. 05, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)
DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2) (Tyrosine kinase-type cell
DE surface receptor HER2) (MLN 19).
GN ERBB2 OR HER2 OR NGL OR NEU.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=6618663; PubMed=3003577;
RA Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,
RA Saito T., Toyoshima K.;
RT "Similarity of protein encoded by the human c-erb-B-2 gene to
RT epidermal growth factor receptor.";
RL Nature 319:230-234(1986).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=86070181; PubMed=299974;
RA Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,
RA McGrath U., Seeburg P.H., Libermann T.A., Schlessinger J.,
RA Francke U., Levinson A., Ullrich A.;
RT "Tyrosine kinase receptor with extensive homology to EGF receptor
RT shares chromosomal location with neu oncogene.";
RL Science 230:1132-1139(1985).
RN [3]
RP SEQUENCE OF 737-1031 FROM N.A.
RX MEDLINE=86016729; PubMed=2995967;
RA Semba K., Kanata N., Toyoshima K., Yamamoto T.;
RT "A v-erbB-related protooncogene, c-erbB-2, is distinct from the
RT c-erbB-1/epidermal growth factor-receptor gene and is amplified in a
RT human salivary gland adenocarcinoma.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).
RN [4]
RP VARIANTS VAL-654 AND VAL-655.
RX MEDLINE=93194196; PubMed=8095488;
RA Ehsani A., Low J., Wallace R.B., Wu A.M.;
RT "Characterization of a new allele of the human ERBB2 gene by allele-
RT specific competition hybridization.";
RL Genomics 15:426-429(1993).
RN [5]
RP FUNCTION. ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.
RN ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A
RN POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
RN ALPHA AND AMPHIREGULIN.
RN -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
RN tyrosine phosphate.
RN -I- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
RN (POTENTIAL).
RN -I- SUBCELLULAR LOCATION: Type I membrane protein.

```

CC -1- PFM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY).  
 CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN  
 CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY  
 CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;  
 CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
 CC at the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: M11767; AAA35808.1; JOINED.  
 DR EMBL: M11761; AAA35808.1; JOINED.  
 DR EMBL: M11762; AAA35808.1; JOINED.  
 DR EMBL: M11763; AAA35808.1; JOINED.  
 DR EMBL: M11764; AAA35808.1; JOINED.  
 DR EMBL: M11765; AAA35808.1; JOINED.  
 DR EMBL: M11766; AAA35808.1; JOINED.  
 DR EMBL: M11730; AAA75493.1; JOINED.  
 DR EMBL: M12036; AAA35978.1; JOINED.  
 DR EMBL: X03363; CAA27060.1; JOINED.  
 DR PIR: A25491; A25491.  
 DR PIR: A24571; A24571.  
 DR HSSP: P11362; 1FGK.  
 DR Genew: HGNC:3430; ERBB2.  
 DR MIM: 164870; ERBB2.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR InterPro: IPR004019; YLP motif.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_pkinase; 1.  
 DR SMART: SM00261; FU; 3.  
 DR SMART: SM00219; TYKc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KW Transmembrane; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KW Polymorphism.  
 FT SIGNAL 1 21 POTENTIAL.  
 FT CHAIN 22 1255 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2.  
 FT DOMAIN 22 652 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 653 675 POTENTIAL.  
 FT DOMAIN 676 1255 CYTOPLASMIC (POTENTIAL).  
 FT NP\_BIND 720 987 PROTEIN KINASE.  
 FT BINDING 726 734 ATP (BY SIMILARITY).  
 FT ACT\_SITE 753 753 ATP (BY SIMILARITY).  
 FT DISULFID 845 845 BY SIMILARITY.  
 FT DISULFID 195 204 BY SIMILARITY.  
 FT DISULFID 199 212 BY SIMILARITY.  
 FT DISULFID 220 227 BY SIMILARITY.  
 FT DISULFID 224 235 BY SIMILARITY.  
 FT DISULFID 236 244 BY SIMILARITY.  
 FT DISULFID 240 252 BY SIMILARITY.  
 FT DISULFID 255 264 BY SIMILARITY.  
 FT DISULFID 266 295 BY SIMILARITY.  
 FT DISULFID 299 311 BY SIMILARITY.  
 FT DISULFID 315 331 BY SIMILARITY.  
 FT DISULFID 334 338 BY SIMILARITY.  
 FT DISULFID 511 520 BY SIMILARITY.  
 FT DISULFID 515 528 BY SIMILARITY.  
 FT DISULFID 531 540 BY SIMILARITY.  
 FT DISULFID 544 560 BY SIMILARITY.

FT DISULFID 563 576 BY SIMILARITY.  
 FT DISULFID 567 584 BY SIMILARITY.  
 FT DISULFID 587 596 BY SIMILARITY.  
 FT DISULFID 600 623 BY SIMILARITY.  
 FT DISULFID 626 634 BY SIMILARITY.  
 FT DISULFID 630 642 BY SIMILARITY.  
 FT MOD\_RES 1139 1139 BY SIMILARITY.  
 FT MOD\_RES 1248 1248 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 124 124 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 187 187 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 259 259 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 530 530 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 571 571 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 629 629 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARIANT 654 654 I -> V.  
 FT VARIANT 655 655 /FTID=VAR\_004077.  
 FT VARIANT 655 655 I -> V.  
 FT CONFLICT 1170 1170 /FTID=VAR\_004078.  
 FT CONFLICT 1255 AA; 137909 MW; 39E9DFDA04DC962 CRC64; P -> A (IN REF. 2).  
 SQ SEQUENCE

Query Match  
 Best Local Similarity 96.94; Score 6602; DB 1; Length 1255;  
 Matches 1221; Conservative 11; Mismatches 18; Indels 10; Gaps 3;

QY 1 MELALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLDMRLHYGCGVQVGNL 60  
 DB 1 MELALCRWGLLALLPFGAASVQCTGDMKRLPASPETHLDMRLHYGCGVQVGNL 60  
 QY 61 ELYLPNTASLSFQDIOEVGVYLIANVQVPLQRLIRVGTQFEDNYALAVLDNG 120  
 DB 61 ELYLPNTASLSFQDIOEVGVYLIANVQVPLQRLIRVGTQFEDNYALAVLDNG 120  
 QY 121 DPLNNTPTVTSASGGLRELQRLTELKGVLIQNPOLCYDITLMKDIPIKNNOLA 180  
 DB 121 DPLNNTPTVTSASGGLRELQRLTELKGVLIQNPOLCYDITLMKDIPIKNNOLA 180  
 QY 121 DPLNNTPTVTSASGGLRELQRLTELKGVLIQNPOLCYDITLMKDIPIKNNOLA 180  
 DB 121 DPLNNTPTVTSASGGLRELQRLTELKGVLIQNPOLCYDITLMKDIPIKNNOLA 180  
 QY 181 LTLIDNRSRACHCSPMCKSRGCSSESDQSLRTVCAGGACRCKGPLPTDCHEQC 240  
 DB 181 LTLIDNRSRACHCSPMCKSRGCSSESDQSLRTVCAGGACRCKGPLPTDCHEQC 240  
 QY 181 LTLIDNRSRACHCSPMCKSRGCSSESDQSLRTVCAGGACRCKGPLPTDCHEQC 240  
 DB 181 LTLIDNRSRACHCSPMCKSRGCSSESDQSLRTVCAGGACRCKGPLPTDCHEQC 240  
 QY 241 AAGCTGPHSDCLALPHNSGICELCPALVTNTTFSMPREGRYTFGASCYVACP 300  
 DB 241 AAGCTGPHSDCLALPHNSGICELCPALVTNTTFSMPREGRYTFGASCYVACP 300  
 QY 241 AAGCTGPHSDCLALPHNSGICELCPALVTNTTFSMPREGRYTFGASCYVACP 300  
 DB 241 AAGCTGPHSDCLALPHNSGICELCPALVTNTTFSMPREGRYTFGASCYVACP 300  
 QY 301 YNYLSTDVGSCTIVCPHNOEVTAEDETQRCCKSKPCARVCYGLGMOYIKANSKFTGIT 360  
 DB 301 YNYLSTDVGSCTIVCPHNOEVTAEDETQRCCKSKPCARVCYGLGMOYIKANSKFTGIT 360  
 QY 301 YNYLSTDVGSCTIVCPHNOEVTAEDETQRCCKSKPCARVCYGLGMOYIKANSKFTGIT 360  
 DB 301 YNYLSTDVGSCTIVCPHNOEVTAEDETQRCCKSKPCARVCYGLGMOYIKANSKFTGIT 360  
 QY 361 ELERAGCKKTFGSLAFIPESFDGDPASNTAPLOEQLOVETLEITGYIISAMPDLP 420  
 DB 361 ELERAGCKKTFGSLAFIPESFDGDPASNTAPLOEQLOVETLEITGYIISAMPDLP 420  
 QY 361 ELERAGCKKTFGSLAFIPESFDGDPASNTAPLOEQLOVETLEITGYIISAMPDLP 420  
 DB 361 ELERAGCKKTFGSLAFIPESFDGDPASNTAPLOEQLOVETLEITGYIISAMPDLP 420  
 QY 421 DLSVFQNLQVIRGILNGAYSLTLQGLISWGLRRLREIGSLALIHNNTHLCFHTV 480  
 DB 421 DLSVFQNLQVIRGILNGAYSLTLQGLISWGLRRLREIGSLALIHNNTHLCFHTV 480  
 QY 421 DLSVFQNLQVIRGILNGAYSLTLQGLISWGLRRLREIGSLALIHNNTHLCFHTV 480  
 DB 421 DLSVFQNLQVIRGILNGAYSLTLQGLISWGLRRLREIGSLALIHNNTHLCFHTV 480  
 QY 481 PMDGLFNNPQALHTANRDEDECVGGLACHOLCARGHGWRGPTQCVNCSQFLRQEC 540  
 DB 481 PMDGLFNNPQALHTANRDEDECVGGLACHOLCARGHGWRGPTQCVNCSQFLRQEC 540  
 QY 481 PMDGLFNNPQALHTANRDEDECVGGLACHOLCARGHGWRGPTQCVNCSQFLRQEC 540  
 DB 481 PMDGLFNNPQALHTANRDEDECVGGLACHOLCARGHGWRGPTQCVNCSQFLRQEC 540  
 QY 541 VEEBRVYQGLPREVYNRHLCPHPEQOPONGSVTCGPBACDCAAHAKDPFVCAR 600  
 DB 541 VEEBRVYQGLPREVYNRHLCPHPEQOPONGSVTCGPBACDCAAHAKDPFVCAR 600  
 QY 541 VEEBRVYQGLPREVYNRHLCPHPEQOPONGSVTCGPBACDCAAHAKDPFVCAR 600  
 DB 541 VEEBRVYQGLPREVYNRHLCPHPEQOPONGSVTCGPBACDCAAHAKDPFVCAR 600  
 QY 601 PSGVKPLSLMPIMKPFDEGACQPCPCINCTHSCVDLDDKGCAPAEGRASLTSIISAVG 660  
 DB 601 PSGVKPLSLMPIMKPFDEGACQPCPCINCTHSCVDLDDKGCAPAEGRASLTSIISAVG 660  
 QY 601 PSGVKPLSLMPIMKPFDEGACQPCPCINCTHSCVDLDDKGCAPAEGRASLTSIISAVG 660  
 DB 601 PSGVKPLSLMPIMKPFDEGACQPCPCINCTHSCVDLDDKGCAPAEGRASLTSIISAVG 660  
 QY 661 ILVVYVGVVFGILIKRQOKIRKYTRRLLOETELVEPLTPSGAMPNQAOKRILKXTEL 720  
 DB 661 ILVVYVGVVFGILIKRQOKIRKYTRRLLOETELVEPLTPSGAMPNQAOKRILKXTEL 720





FT DISULFID 569 586 BY SIMILARITY.  
 FT DISULFID 589 598 BY SIMILARITY.  
 FT DISULFID 602 625 BY SIMILARITY.  
 FT DISULFID 628 636 BY SIMILARITY.  
 FT DISULFID 632 644 BY SIMILARITY.  
 FT MOD\_RES 1141 1141 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1250 1250 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 68 68 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 188 188 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 260 260 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 532 532 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 573 573 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 631 631 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT VARIANT 661 661 V -> E (IN ONCOGENIC NEO).  
 SQ SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;

Query Match 85.3%; Score 581; DB 1; Length 1257;  
 Best Local Similarity 85.3%; Pred. No. 6.6e-304;  
 Matches 1077; Conservative 58; Mismatches 115; Indels 12; Gaps 5;

QY 1 MELAALCRWGLLLALPPGAASCTGCTGDMKRLPASPEHLDMLRHLYOGCCVQVGNL 60  
 DB 1 MELAAMCRWGLLLALPPGIAGCTGCTGDMKRLPASPEHLDMLRHLYOGCCVQVGNL 60  
 QY 61 ELTYLPTNASLSPLODIOEVGVYLIANOVROVPLRLTYRGTOLEFENYALALVDNG 120  
 DB 61 ELTYLPTNASLSPLODIOEVGVYLIANOVROVPLRLTYRGTOLEFENYALALVDNR 120  
 QY 121 DPLNNTTPTVT-GASPGGLRELQRLSLTEILKGVYLIQNPOLCYQDTILWKDIFFHKNNQL 179  
 DB 121 DPLNNTTPTVT-GASPGGLRELQRLSLTEILKGVYLIQNPOLCYQDMVWLVKWDYFRKNNQL 180  
 QY 121 DPONNVAASSTPGRIPEGLRELQRLSLTEILKGVYLIQNPOLCYQDMVWLVKWDYFRKNNQL 180  
 DB 180 ALTLIDTRSRACHPSCPMCKSGKSCWESSSDCQLRTYVAGGACAKGELPDDCCEQ 239  
 QY 181 APVDIDTRSRACHPSCPMCKSGKSCWESSSDCQLRTYVAGGACAKGELPDDCCEQ 240  
 DB 181 APVDIDTRSRACHPSCPMCKSGKSCWESSSDCQLRTYVAGGACAKGELPDDCCEQ 240  
 QY 240 CAAGCTGKXSDCLACILFHNHSGICELHCPALVYNTDFESMNPREGRTFGASCYTAC 299  
 DB 241 CAAGCTGKXSDCLACILFHNHSGICELHCPALVYNTDFESMNPREGRTFGASCYTTC 300  
 QY 300 PYNLTSTDVGSCTLVCPILHNOEVTAEEDGTORCEKSPCARVCGYGLMOKYIKANSKEIGI 359  
 DB 301 PYNLTSTEVSCTLVCPILHNOEVTAEEDGTORCEKSPCARVCGYGLMOKYIKANSKEIGI 360  
 QY 360 TELEPAGCKKIFGSLAFIPESFDDPASNAPLOPELOVETLEITGLYLSAMPDGL 419  
 DB 361 NVQEPDCKKIFGSLAFIPESFDDPASNAPLOPELOVETLEITGLYLSAMPDGL 420  
 QY 420 PDLISFQNLQYIRGRILHNGAYSJTLQGLGHSWLGRLSELGSGALLIHNHNLCPVHT 479  
 DB 421 RDLISFQNLQYIRGRILHNGAYSJTLQGLGHSWLGRLSELGSGALLIHNHNLCPVHT 480  
 QY 480 VPMQOLFRRPHQALLHNTANRPEDE-CVGEGLACIQLCARHCHWCPGTQCCVNSQPIRGQ 538  
 DB 481 VPMQOLFRRPHQALLHNTANRPEDE-CVGEGLACIQLCARHCHWCPGTQCCVNSQPIRGQ 540  
 QY 539 ECVEECRVLQGLPREYVNAHRCPLCPHECQFQNSVTCFGEADQCVACAHYKDPRECV 598  
 DB 541 ECVEECRVLQGLPREYVNAHRCPLCPHECQFQNSVTCFGEADQCVACAHYKDPRECV 600  
 QY 599 RCPGSGVPRDLSYMIKMFPEDEGACQPCPINCCHSCVDLDDKGPAGQASPLTSYSAV 658  
 DB 601 RCPGSGVPRDLSYMIKMFPEDEGACQPCPINCCHSCVDLDDKGPAGQASPLTSYSAV 660  
 QY 659 VGLILVVVLGVFGLILKRRQKIRKXTMRLLQETLVEPLTPSGAMPQAOBRILKET 718  
 DB 661 VGLILVVVLGVFGLILKRRQKIRKXTMRLLQETLVEPLTPSGAMPQAOBRILKET 720  
 QY 719 ELRKVKYLGSGAFNNFTVSWFLRV-----KVSASHLEVLRENTSPKANKELDEAYV 773  
 DB 721 ELRKVKYLGSGAFNNFTVSWFLRV-----KVSASHLEVLRENTSPKANKELDEAYV 775  
 QY 774 MAGVGSPIVSRLLIGICTSTVQVLTQLMPYGCCLDHNRENRGRGLSGDGLNWMQIAKCM 833

DB 776 MAGVGSPIVSRLLIGICTSTVQVLTQLMPYGCCLDHNRENRGRGLSGDGLNWMQIAKCM 835  
 QY 834 STLEDRLVHRRLAARNTVYKSPNVTITDPEGLALLDIDETENYADGKVYIKMALES 893  
 DB 836 STLEDRLVHRRLAARNTVYKSPNVTITDPEGLALLDIDETENYADGKVYIKMALES 895  
 QY 894 ILRRRTTHSDWASVGVTVWELMTFGAKPYDGI PAREIPDLLEKSRLLPQPICTIDVYM 953  
 DB 896 ILRRRTTHSDWASVGVTVWELMTFGAKPYDGI PAREIPDLLEKSRLLPQPICTIDVYM 955  
 QY 954 IMVCKMIDSECRPRELIVSEFSRMARDPQFVVIQNEEDCPASPLDSTFYRSLLEDD 1013  
 DB 956 IMVCKMIDSECRPRELIVSEFSRMARDPQFVVIQNEEDCPASPLDSTFYRSLLEDD 1015  
 QY 1014 MGLDYAEEYLVPOGCFPCDPAPAGGVVHRRHSSSTRSGGGLTGLBPSSEBAPS 1073  
 DB 1016 MGLDYAEEYLVPOGCFPCDPAPAGGVVHRRHSSSTRSGGGLTGLBPSSEBAPS 1075  
 QY 1074 PLAPSEGASDVFDGDLGMAAKGLQSLPTHDPSFLORYSEDPTVPLPSETDGYVAPLTC 1133  
 DB 1076 PLAPSEGASDVFDGDLGMAAKGLQSLPTHDPSFLORYSEDPTVPLPSETDGYVAPLTC 1135  
 QY 1134 SPQPEYVNOSEVQPPPLTPESGLPVPVPAAGTLERPKLTSKGKGVNDVAFGAVEN 1193  
 DB 1136 SPQPEYVNOSEVQPPPLTPESGLPVPVPAAGTLERPKLTSKGKGVNDVAFGAVEN 1195  
 QY 1194 PEYLTFQGAAPQPPHPPAFSPAFDNLVYDODPPEERGAPSTFTFGTPTAENPEVGLDV 1253  
 DB 1196 PEYLTFQGAAPQPPHPPAFSPAFDNLVYDODPPEERGAPSTFTFGTPTAENPEVGLDV 1255  
 QY 1254 PV 1255  
 DB 1256 PV 1257

RESULT 3  
 ERB2\_MESAU STANDARD; PRT; 1254 AA.  
 ID ERB2\_MESAU  
 AC 060553;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-2 precursor (EC 2.7.1.112)  
 DE (p185erbB2) (NEU proto-oncogene) (C-erbB-2).  
 GN ERB2 OR NEU.  
 OS Mesocricetus auratus (Golden hamster).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
 OC Mesocricetus.  
 RN NCBI\_Taxid=10036;  
 RX (1)  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Nerve;  
 RX MEDLINE=94193007; PubMed=7908275;  
 RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,  
 RA Yamazaki Y., Ishikawa T.;  
 RT "Cloning and activation of the Syrian hamster neu proto-oncogene";  
 RL Gene 140:251-255(1994).  
 CC -1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX.  
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP130 IS A  
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-  
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration



QY 1076 APSEGASDVFDGLGMAKGLQSLPTDHPSPLOFQSESDPTVPLPSETDGVAPLTCSP 1135  
 DB 1076 APSEGASDVFESEGLMGATKQSPQISPRDLSPLOQSYSDPTLPPLTERIDGVAPLACSP 1135  
 QY 1136 QPEYVNOQDVPDPSPSPREBPPLPAAPACATLERRATLSPGKGVGVKVFANAGAVENPE 1195  
 DB 1136 QPEYVNOQDVPDPSPSPREBPPLPAAPACATLERRATLSPGKGVGVKVFANAGAVENPE 1195  
 QY 1196 YLTPGGAAPQPPHPPAFSPAFNLTYYMDODPPERGAAPSTFKGPTTAENPEYLGIDVPV 1255  
 DB 1196 YLTPGGAAPQPPHPPAFSPAFNLTYYMDODPPERGAAPSTFKGPTTAENPEYLGIDVPV 1255

RESULT 4  
 EGFR\_HUMAN STANDARD; PRT: 1210 AA.  
 ID EGFR\_HUMAN P05533; P05626; Q14225; Q9UMD7; Q9UMD8; Q9UMG5; Q92795; Q00732;  
 AC Q00688; Q9B2S2; Q9H2C9; Q9GZL1; Q9H3C9;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-NOV-1997 (Rel. 35, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Epidermal growth factor receptor precursor (BC 2.7.1.112) (Receptor  
 protein-tyrosine kinase ErbB-1).  
 GN EGFR OR ERBB1.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=84219729; PubMed=6328312;  
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,  
 Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,  
 Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;  
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant  
 expression of the amplified gene in A431 epidermoid carcinoma cells";  
 RL Nature 309:418-425(1984).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RA TISSUE=Placenta; PubMed=7654366;  
 RA Ilekis J.V., Scarf B.C., Scoccia B.;  
 RT "Possible role of variant RNA transcripts in the regulation of  
 RT epidermal growth factor receptor expression in human placenta";  
 RL Mol. Reprod. Dev. 41:149-156(1995).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RA TISSUE=Placenta; PubMed=8918811;  
 RA Reiter J.L., Mahle N.J.;  
 RT "A 1.8 kb alternative transcript from the human epidermal growth  
 RT factor receptor gene encodes a truncated form of the receptor";  
 RL Nucleic Acids Res. 24:4050-4056(1996).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RX TISSUE=Placenta;  
 RA TISSUE=Placenta; PubMed=9103389;  
 RA Ilekis J.V., Garita J., Niederberger C., Scoccia B.;  
 RT "Expression of a truncated epidermal growth factor receptor-like  
 RT protein (TEGFR) in ovarian cancer";  
 RL Gynecol. Oncol. 65:36-41(1997).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORMS 3 AND 4).  
 RX TISSUE=Placenta;  
 RA MEDLINE=11100872; PubMed=1161793;  
 RA Reiter J.L., Threadgill D.W., Riley G.D., Strunk K.E., Danielsen A.J.,  
 RA Schehl Sinclair C., Pearse R.S., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramanian S., Crosley T.D., Magnusen T.R., James C.D.,  
 RA Mainle N.J.;  
 RT "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative EGFR transcripts encoding truncated receptor  
 RT isoforms";  
 RL Genomics 71:1-20(2001).

RN [6]  
 RP SEQUENCE OF 575-687 FROM N.A.  
 RA Reiter J.L., Threadgill D.W., Danielsen A.J., Schehl C.M.,  
 RA Lampland A.L., Balasubramanian S., Crosley T.O., Magnusen T.R.,  
 RA Mainle N.J.;  
 RT "Human and mouse alternative EGFR transcripts encoding only the  
 RT extracellular domain of the receptor";  
 RL Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.  
 RN [7]  
 RP SEQUENCE OF 713-924 FROM N.A.  
 RX MEDLINE=84196372; PubMed=6326261;  
 RA Lin C.R., Chen W.S., Krueger M., Stolarsky L.S., Weber W.,  
 RA Evans R.M., Verma I.M., Gill G.N., Rosenfield M.G.;  
 RT "Expression cloning of human EGFR receptor complementary DNA: gene  
 RT amplification and three related messenger RNA products in A431  
 RT cells";  
 RL Science 224:843-848(1984).  
 RN [8]  
 RP SEQUENCE OF 150-962 FROM N.A.  
 RX MEDLINE=84245835; PubMed=6330563;  
 RA Xu Y.H., Ishii S., Clark A.J.L., Sullivan M., Wilson R.K., Ma D.P.,  
 RA Roe B.A., Merlino G.T., Pastan I.;  
 RT "Human epidermal growth factor receptor cDNA is homologous to a  
 RT variety of RNAs overproduced in A431 carcinoma cells";  
 RL Nature 309:806-810(1984).  
 RN [9]  
 RP SEQUENCE OF 1028-1210 FROM N.A.  
 RX MEDLINE=85046483; PubMed=6093780;  
 RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,  
 RA O'Malley B.W.;  
 RT "Isolation of an evolutionarily conserved epidermal growth factor  
 RT receptor cDNA from human A431 carcinoma cells";  
 RL Biochem. Biophys. Res. Commun. 124:125-132(1984).  
 RN [10]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=88217333; PubMed=3329716;  
 RA Haley J.D., Whittle N., Bennett P., Kinsington D., Ullrich A.,  
 RT "The human EGF receptor gene: structure of the 110 kb locus and  
 RT identification of sequences regulating its transcription";  
 RL Oncogene Res. 1:375-396(1987).  
 RN [11]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=91107677; PubMed=1988448;  
 RA Haley J.D., Waterfield M.D.;  
 RT "Contributory effects of de novo transcription and premature  
 RT transcript termination in the regulation of human epidermal growth  
 RT factor receptor proto-oncogene RNA synthesis";  
 RL J. Biol. Chem. 266:1746-1753(1991).  
 RN [12]  
 RP SEQUENCE OF 1-29 FROM N.A.  
 RX MEDLINE=85270438; PubMed=2391899;  
 RA Ishii S., Xu Y.H., Stratton R.H., Roe B.A., Merlino G.T., Pastan I.;  
 RT "Characterization and sequence of the promoter region of the human  
 RT epidermal growth factor receptor gene";  
 RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).  
 RN [13]  
 RP SEQUENCE OF 540.  
 RA Kohda D.;  
 RT Submitted (SEP-1997) to the SWISS-PROT data bank.  
 RN [14]  
 RP RECEPTOR ACTIVITY.  
 RX MEDLINE=84191554; PubMed=6325948;  
 RA Mroczkowski B., Mosig G., Cohen S.;  
 RT "ATP-stimulated interaction between epidermal growth factor receptor  
 RT and supercoiled DNA";  
 RL Nature 309:270-273(1984).  
 RN [15]  
 RP PHOSPHORYLATION.  
 RX MEDLINE=89278137; PubMed=2543678;  
 RA Margolis B.L., Lax I., Kris R., Domagalian M., Honegger A.M.,  
 RA Hawk R., Givol D., Ullrich A., Schlessinger J.;  
 RT "All autophosphorylation sites of epidermal growth factor (EGF)

RT receptor and HER2/neu are located in their carboxyl-terminal tails.  
 RT Identification of a novel site in EGF receptor."  
 RL J. Biol. Chem. 264:10667-10671 (1989).  
 RN [16]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-128; ASN-175; ASN-413; ASN-444 AND  
 RP ASN-528.  
 RX MEDLINE=96398132; Pubmed=8962717;  
 RA Smith K.D., Davies M.J., Bailey D., Renouf D.V., Hounsell E.F.;  
 RT "Analysis of the glycosylation patterns of the extracellular domain of  
 RT the epidermal growth factor receptor expressed in Chinese hamster  
 RT ovary fibroblasts."  
 RL Growth Factors 13:121-132 (1996).  
 RN [17]  
 RP CARBOHYDRATE-LINKAGE SITES ASN-56; ASN-352; ASN-361; ASN-568 AND  
 RP ASN-603.  
 RX MEDLINE=20198209; Pubmed=10721668;  
 RA Sato C., Kim J.-H., Abe Y., Saito K., Yokoyama S., Kohda D.;  
 RT "Characterization of the N-oligosaccharides attached to the atypical  
 RT Asn-X-Cys sequence of recombinant human epidermal growth factor  
 RT receptor."  
 RL J. Biochem. 127:65-72 (2000).  
 RN [18]  
 RP PARTIAL SEQUENCE, AND DISULFIDE BONDS.  
 RX MEDLINE=98225196; Pubmed=9556602;  
 RA Abe Y., Odaka M., Inagaki F., Iax I., Schlessinger J., Kohda D.;  
 RT "Disulfide bond structure of human epidermal growth factor receptor."  
 RL J. Biol. Chem. 273:11150-11157 (1998).  
 RN [19]  
 RP REVIEW.  
 RX MEDLINE=87297456; Pubmed=3039909;  
 RA Carpenter G.;  
 RT "Receptors for epidermal growth factor and other polypeptide  
 RT mitogens."  
 RL Annu. Rev. Biochem. 56:881-914 (1987).  
 CC -1- FUNCTION: Receptor for EGF, but also for other members of the EGF  
 CC family, as TGF-alpha, amphiregulin, betacellulin, heparin-binding  
 CC EGF-like growth factor, GP30 and vaccinia virus growth factor. Is  
 CC involved in the control of cell growth and differentiation.  
 CC -1- FUNCTION: Isoform 2/truncated isoform may act as an antagonist.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Isoform 2 is  
 CC secreted.  
 CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; 1/p170 (shown here), 2/p60/  
 CC truncated isoform/TEGFR, 3/p110 and 4; are produced by  
 CC alternative splicing.  
 CC -1- TISSUE SPECIFICITY: Expressed in placenta. Isoform 2 is also  
 CC expressed in ovarian cancers.  
 CC -1- MISCELLANEOUS: Binding of EGF to the receptor leads to  
 CC dimerization, internalization of the EGF-receptor complex,  
 CC induction of the tyrosine kinase activity, stimulation of cell DNA  
 CC synthesis, and cell proliferation.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: X00588; CA28240.1; -  
 DR EMBL: U95089; AAB53063.1; -  
 DR EMBL: U48722; AAC50802.1; -  
 DR EMBL: U48723; AAC50804.1; -  
 DR EMBL: U48724; AAC50796.1; -  
 DR EMBL: U48725; AAC50797.1; -  
 DR EMBL: U48726; AAC50798.1; -

Query Match 44.5%; Score 3033; DB 1; Length 1210;  
 Best Local Similarity 48.7%; Pired. No. 3; 7e-155;  
 Matches 619; Conservative 175; Mismatches 360; Indels 116; Gaps 25;

QY	11 LLLALLPGAA--STQVCTGTDMLRLPASPTHLDMRLHLYQGCGVQVGNLELYLPTN 68	QY	308 VGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGITLE-FA 366
DB	14 LLAALCPASRLAEKKVCGGTSTNKLTLQGTGFEDHFLSLQRMENNECVTLGNETIVQGN 73	DB	304 HGSQVRACGADSYEM-EEDGVYKCKCKGCGPKVCNGIGIFBK-DSLSINATNFKHFN 361
QY	69 ASLSPLQDIQEVQVYLLAHNQVQVPLQRLRIVNGTQGFEDNNYALAVLNDPPLNTTP 128	QY	367 CKKIGSLAFIPESFDGPASTAFLQREQLQVFTLEITGYLYISAMPDLPLSLVQ 426
DB	74 YLSPLKTLQEVAGVYLLALNTVERIPLENLIQINGNNYENSLAVLAVLSND----- 126	DB	362 CTISGDDHLIPVAERGDSFHTPLDQLQEDLILKTVKELTGFLIQAMPENRTDLHAF 421
QY	129 VTGASPGGLREQLRLSLTEILKGVLIQRNPQCYQDTILMDIFHNKQLALTIDTR 188	QY	427 NLQVIRGRILHNGAVSLTLQGLISLWLGRLSRLRELSGALILHNTHLCFVTVPMDOLE 486
DB	127 ---AKTGKLELPMNLQELHGAIRFSNNPPLCNVBSIQMRDIISSDPLSMMSMDFOH 183	DB	422 NLEIRGRTHQHQGSLAVSLNITSLGRSLKESDDVILSGKNLCYANTIMKKLIF 481
QY	189 SPACPCSPMCKSGSCWGESSEDCSLTRTVACGCA-RCKGPLPTDCHECCAACTGP 247	QY	487 RNPHQALLHTANRPDEVCVEGLACHQLCARHCKGPGPPTQCVNCSQFLRQGEVCEGRV 546
DB	184 LSCCKCPSCPCNCSWCGAGENCCOKTLKICAOCCSGRCRCKSPSDCHNCAAGCTGP 243	DB	482 GTSGCKTKIISNRGNSCKATQGVCHALCSBEGCPEPRDCCVSRNRSRGEVCDKNL 541
QY	248 KASDCLACLFHNSGICELCPALVTYNTDPESMPNDEGRRTFASACTAPRYVLSLD 307	QY	547 LQGLREYVNAHCLPCPECPQNGSTCCGPEADQCVACHYNDPPFCVARGCSGVKP 606
DB	244 RESDGLVCKRFDEATCKDTCPLPLVPTTYQMVDNBEKISFQATCYKCPRYNVTD 303	DB	542 LEGERREVENSECTIQCPBELPQMMNITLTCRGCDNCTIQCHYTDGPHCVAGVWG 601
QY	308 VGSCTLVCPLNQEVTAEDGTORCEKSKPCARVCYGLGMOYIKANSKFIGITLE-FA 366	QY	607 DLSTYPIWKFPDEBACQPCPILNTHSCVDDLDKCPAORASPLTISVAVG---ILL 663
DB	304 HGSQVRACGADSYEM-EEDGVYKCKCKGCGPKVCNGIGIFBK-DSLSINATNFKHFN 361	DB	602 ENNTL-VMKYADAGVCHLCHPNCTYGTGGLGECPLNGKIP--SLATGVVGAALLLL 658
QY	367 CKKIGSLAFIPESFDGPASTAFLQREQLQVFTLEITGYLYISAMPDLPLSLVQ 426	QY	664 VVVLGVVGLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAOIRLIKETELKV 723
DB	362 CTISGDDHLIPVAERGDSFHTPLDQLQEDLILKTVKELTGFLIQAMPENRTDLHAF 421	DB	659 VVALGIG--LFMRRHIVRKRTLRLQERELVEPLTPSGEAPAPQALLRIKETEFPKI 715
QY	427 NLQVIRGRILHNGAVSLTLQGLISLWLGRLSRLRELSGALILHNTHLCFVTVPMDOLE 486	QY	724 KYLSGAGFGFNNFTVSFMLRVPKVSASHLEV---LRENTSPKANKETLDEAYVWAGVS 779
DB	422 NLEIRGRTHQHQGSLAVSLNITSLGRSLKESDDVILSGKNLCYANTIMKKLIF 481	DB	716 KYLSGAGFG---TYKGLMIPBEGKVKIPAIKELREATSPKANKETLDEAYVWASVDN 771
QY	487 RNPHQALLHTANRPDEVCVEGLACHQLCARHCKGPGPPTQCVNCSQFLRQGEVCEGRV 546	QY	780 PYVSRLLGICLTSTVQVLTQMLPYGCLLDHVRNREGRLGQDLLNMCQIAKAGSYLSDV 839
DB	482 GTSGCKTKIISNRGNSCKATQGVCHALCSBEGCPEPRDCCVSRNRSRGEVCDKNL 541	DB	772 PHVCLLGICLTSTVQVLTQMLPYGCLLDHVRNREGRLGQDLLNMCQIAKAGMYLEDR 831
QY	547 LQGLREYVNAHCLPCPECPQNGSTCCGPEADQCVACHYNDPPFCVARGCSGVKP 606	QY	840 RLVRHDLAARVNVKVSPPHVKITDPCGLARLLDIETEVHAGGVKIPFMMLLESTLRRF 899
DB	542 LEGERREVENSECTIQCPBELPQMMNITLTCRGCDNCTIQCHYTDGPHCVAGVWG 601	DB	832 RLVRHDLAARVNVKIPHVKITDPCGLARLLDIETEVHAGGVKIPFMMLLESTLRRF 891
QY	607 DLSTYPIWKFPDEBACQPCPILNTHSCVDDLDKCPAORASPLTISVAVG---ILL 663	QY	900 THQSDVMSYGVTVMLMTFGAKPYDGIIPAREIPDLLEKGERLPPPICTIDVVMIMVNCW 959
DB	602 ENNTL-VMKYADAGVCHLCHPNCTYGTGGLGECPLNGKIP--SLATGVVGAALLLL 658	DB	892 THQSDVMSYGVTVMLMTFGSKPYDGIIPASISSTILEKGERLPPPICTIDVVMIMVNCW 951
QY	664 VVVLGVVGLIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAOIRLIKETELKV 723	QY	960 MIDSCRRPREFELVSEFRRMARQRFVILQ-NEDLPGASPLDSTFTFRLLEDMDGDLV 1018
DB	659 VVALGIG--LFMRRHIVRKRTLRLQERELVEPLTPSGEAPAPQALLRIKETEFPKI 715	DB	952 MIDASRRKREFELVSEFRRMARQRFVILQVIGDERMLPSPITDSNFYALMDEEDMDVDV 1011
QY	724 KYLSGAGFGFNNFTVSFMLRVPKVSASHLEV---LRENTSPKANKETLDEAYVWAGVS 779	QY	1019 DAEEYLVQCGFFCPDPAPAGAGVWHHRHRSSTRSGGDLTLGLEPSEEBEAPSPAPLS 1078
DB	716 KYLSGAGFG---TYKGLMIPBEGKVKIPAIKELREATSPKANKETLDEAYVWASVDN 771	DB	1012 DAEEYLVQCGFF-----SSPSTSRTPLLSS 1037



FT DISULFID 620 628 BY SIMILARITY.  
 FT DISULFID 624 636 BY SIMILARITY.  
 FT MOD\_RES 680 680 PHOSPHORYLATION (BY PKC) (BY SIMILARITY).  
 FT MOD\_RES 1092 1092 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1110 1110 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1172 1172 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1197 1197 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 128 128 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 175 175 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 196 196 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 352 352 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 413 413 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 444 444 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 528 528 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 568 568 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 603 603 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 623 623 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CONFLICT 19 19 C -> S (IN REF. 2).  
 FT CONFLICT 539 539 C -> W (IN REF. 5).  
 FT CONFLICT 991 991 L -> F (IN REF. 4).  
 FT CONFLICT 1116 1117 HP -> DR (IN REF. 6).  
 SQ SEQUENCE 1210 AA; 134853 MW; 690E20D46DFD22F5 CRC64;

Query Match 44.1%; Score 3007; DB 1; Length 1210;  
 Best Local Similarity 48.6%; Pred. No. 9e-154;  
 Matches 621; Conservative 169; Mismatches 367; Indels 120; Gaps 27;

QY 11 LLLALLPPGAA-STOVCTGTDMKRLPASPEHLMRLYOGCQVQGNLELYLPTN 68  
 DB 14 LTRALCAAGALEKXKVCQGTNSRLTGLTFEDHFLSLQRMNNECVLNLLETIVORN 73  
 QY 69 ASLSFLDIDIOVQYVLIANNQVQVLOGLRITRGQLFEDYALALVINDGPNLNTT 128  
 DB 74 YDSFLKTIQEVAVYLIANTVERIPLEMLQIIRGALYENTYALALSN----- 124  
 QY 129 VTGASPGRLRELRLSTLTELKGGVLIQRPOLCYOQTLIMKDI---FRKNQLALTLI 184  
 DB 125 -YGNRTGLRELPRNLIQELLIGAVRFSNNPILCMOTIQRDIYQVNFSSMSMDL--- 180  
 QY 185 DTNRBRACHPCSPCKSRCKSGESSEDCOSLTRVCGGCA-RCKGLPTDCCHEQCAAG 243  
 DB 181 -QSHSSCPKCDPCSPGSCWGGEENCQKLTIKICQOCSHRCRGRSPDCCCHNOCAAG 239  
 QY 244 CTGPKHSDCLACHFNHSGICELHCPALVTYNTDTPFSPMNEPGRVFGASCYACPYNY 303  
 DB 240 CTGPRESDCLVCCQFQDEATCKTCPCPLMLYNTPTTYMDVNPBCKXSFGATCYKCEPRNY 299  
 QY 304 LSTDVSGCTLVCPPLHNOVTAEADGTQRCCKSKPCACVYCYGLQMYIKANSKFIITTELE 363  
 DB 300 VVTDHSGCVACGDEYEV-EDDGIKCKKCDGFCRVCNGIGIGEFK-DTLSINATNIK 357  
 QY 364 -FACCKKIFGSLAPESFDGDPASNTAPLOPELOVFELEETGLYISAPDSLPDL 422  
 DB 358 HFKCTAISGDHLPLVAFKSDSTRTPPDPRELETLKTVKELTGLLQAMPDMWTDL 417  
 QY 423 SVFQNLQVIRGRILHNGAYSLTLOGLSIWLGRSLRELSSGLALIHNTHLGFVHTVPW 482  
 DB 418 HAFENLEIRGRITQHGOFSLAVGLNTSLGRSLKEIDGVDYIISGNRLCYANTINW 477  
 QY 483 DQLFNRPHQALHNPPEDECVGEGLAQHLCARSGCNRPPOCNCQGFIRGQECVE 542  
 DB 478 KKLFGTNGKTKIMNNAEKDCQAVNHVNCPLSSBECWGPEDDCVSCNVSRGRCVSE 537  
 QY 543 ECRVQLPREYVNAARLCPHCEPCQFQNGSVTCFGEADQCAVCAHYKPPFCVAPRPS 602  
 DB 538 KCNLTBEPREFEVNSCICHPRECLPQAMNITLTGSPNCIQCAHYIDGPHCVKTCRA 597  
 QY 603 GVKPDLSTYPIWKPPDEGACQPCPINCSTCYDLDKGPAPORASPLTISAVVIGIL 662  
 DB 598 GIMGENTNL-VMKYADANNVCHLHANCATGCGAPGIGQCEVWPSGPKIPISATIGVIGL 656  
 QY 663 LVVVLGVVFGI-LIKRQOKIRKRYTMRLLQETELVPLTPSGAMPNQAMRLIKETELR 721

DB 657 LPIV-VALGIGLFMRRIHVKRLRLRLDERELVEBLTPSGEAPNQAHLILETERK 715  
 QY 722 KKVYSGAGFENFTVSMLEVPKVSASHLEV---LRENTSPANKREILDEAVYMAV 777  
 DB 716 KKVYSGAGF---TVYKGLMTPBEGEKYKIPVALKELREARSPANKREILDEAVYMAV 771  
 QY 778 GSPVYSRLIGICLTSTVOLVTQMLPYGCLLDHVRNRGRLSQDILLNMCQIAKMSYLE 837  
 DB 772 DHPHVCRLIGICLTSTVOLITQMLPYGCLLDHVRNRGRLSQDILLNMCQIAKMSYLE 831  
 QY 838 DRLVLRDLAARNVTVKSPNHYKIDPGLAFLDIDEMRYADGKVPKMALESILHR 897  
 DB 832 DRLVLRDLAARNVTVKSPNHYKIDPGLAFLDIDEMRYADGKVPKMALESILHR 891  
 QY 898 RFTTHSDVMSYGVTVWELMTFGAKPEYDGIIPAREIDLEKGERLPQPPICITDYVYIMWK 957  
 DB 892 IYTHQSDVMSYGVTVWELMTFGSKPYDGIIPAREIDLEKGERLPQPPICITDYVYIMWK 951  
 QY 958 CMMIDSECPRERELVESESRKAPDPORFVYQ-NEDGAPSPULSTPYRSLLEDMDKD 1016  
 DB 952 CMMIDSECPRERELVESESRKAPDPORFVYQ-NEDGAPSPULSTPYRSLLEDMDKD 1011  
 QY 1017 LVDAEYVLPQOQFCPPDPAPGAGVHRRSSSTRSGGGLTLGLEPSEEARPSPLA 1076  
 DB 1012 VVDADDEYLIPQOQF-----NSPST-----SRTPLL 1037  
 QY 1077 PSEGAGSDVFDGDLGMAAGLIQSIPTHDPSFLQRYSEDPIVLPSET-DGYVAPLTCS 1134  
 DB 1038 SLSLATSN---NSTVACINRNGSCRVKEDAPLQRYSSDPGATVEDNIDDAFL----- 1087  
 QY 1135 PQPEYVNPDPVPPQSPREPEGLPAPAPGATLEBAKTLSPKNGVNVDFVFGAGAVENP 1194  
 DB 1088 PVPEYVNP-SVPRKAPGAVQNPVHNQPLHP-----APGRDLHYQN-THSNVAGVP 1136  
 QY 1195 EYL-TPQGAAPQPPPPAPFSPAFNLYWMDQ-----DP-----PERGAPSPSTF 1237  
 DB 1137 EYLNTAQ-----PCLSSGFSNPSALMIQKSHQMSLDNDVDQOQDFPRETPNGIF 1187  
 QY 1238 KGTPTAENPEYGLDVP 1254  
 DB 1188 KG-PTAENAYLVRVP 1203  
 RESULT 6  
 ERB4\_HUMAN  
 ID ERB4\_HUMAN STANDARD; PRT; 1308 AA.  
 AC Q15303;  
 DT 15-DEC-1998 (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112)  
 DE (p180erbB4) (Tyrosine kinase-tyrosine cell surface receptor HER4).  
 GN ERB4 OR HER4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM JM-A).  
 RC TISSUE=breast carcinoma;  
 RX MEDLINE=93189574; PubMed=8383326;  
 RA Plozman G.D., Culoucou J.-M., Whitney G.S., Green J.M., Carlton G.W.,  
 Foy L., Neubauer M.G., Shoyab M.,  
 RT "Ligand-specific activation of HER4/p180erbB4, a fourth member of the  
 RT epidermal growth factor receptor family".  
 RL Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750 (1993).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM JM-A AND JM-B).  
 RC TISSUE=fetal brain;  
 RX MEDLINE=97476287; PubMed=9334263;  
 RA Elenius K., Cortes G., Paul S., Choi C.J., Rio C., Plozman G.D.,  
 RA Klagesbrun M.;

"A novel juxtamembrane domain isoform of HER4/ErB4. Isoform-specific tissue distribution and differential processing in response to phorbol ester".  
 J. Biol. Chem. 272:26761-26768(1997).  
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NRG-2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND NTRK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.  
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.  
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B; ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.  
 CC -1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART, KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM, PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS, LUNG, SALIVARY GLAND, AND PANCREAS.  
 CC -1- PPM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC EMBL: L07868; AAB59446.1; -  
 DR HSSP: P1362; IFGK.  
 DR Gene: HGNC:3432; ERBB4.  
 DR MIM: 600543; -  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_Pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_Pkinase.  
 DR InterPro: IPR004019; YLP\_motif.  
 DR Pfam: PF00063; Pkinase\_1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR Pfam: PF02757; YLP; 2.  
 DR ProDom: PD000001; Euk\_Pkinase; 1.  
 DR SMART: SM00261; FU; 4.  
 DR SMART: SM00219; Tyrc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00109; PROTEIN\_KINASE\_TYR; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;  
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;  
 KM Alternative splicing.  
 FT SIGNAL 1 25  
 FT CHAIN 26 1308  
 FT DOMAIN 26 651  
 FT TRANSMEM 652 675  
 FT DOMAIN 676 1308  
 FT DOMAIN 186 334  
 FT DOMAIN 496 633  
 FT DOMAIN 718 985  
 FT NP\_BIND 724 732  
 FT BINDING 751 751  
 FT ACI\_SITE 843 843  
 FT DISULFID 189 197  
 FT DISULFID 193 205  
 FT DISULFID 213 221  
 FT DISULFID 217 229  
 FT DISULFID 230 238

FT DISULFID 234 246 BY SIMILARITY.  
 FT DISULFID 249 259 BY SIMILARITY.  
 FT DISULFID 262 289 BY SIMILARITY.  
 FT DISULFID 293 304 BY SIMILARITY.  
 FT DISULFID 308 323 BY SIMILARITY.  
 FT DISULFID 326 330 BY SIMILARITY.  
 FT DISULFID 503 512 BY SIMILARITY.  
 FT DISULFID 507 520 BY SIMILARITY.  
 FT DISULFID 523 532 BY SIMILARITY.  
 FT DISULFID 536 552 BY SIMILARITY.  
 FT DISULFID 555 569 BY SIMILARITY.  
 FT DISULFID 559 577 BY SIMILARITY.  
 FT DISULFID 580 589 BY SIMILARITY.  
 FT DISULFID 593 614 BY SIMILARITY.  
 FT DISULFID 617 625 BY SIMILARITY.  
 FT DISULFID 621 633 BY SIMILARITY.  
 FT MOD\_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 181 181 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 253 253 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 410 410 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 473 473 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 548 548 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC...) (POTENTIAL).  
 FT VARSPLIC 626 648 NGPISHDCIYFWTGHSTLPQHA -> IGSSIEDICGLMD (IN ISOFORM JM-B).  
 SQ SEQUENCE 1308 AA; 146807 MW; 5E4AE80985D8761 CRC64;  
 Query Match 41.9%; Score 2852.5; DB 1; Length 1308;  
 Best Local Similarity 44.1%; Pred. No. 18e-145;  
 Matches 596; Conservative 187; Mismatches 385; Indels 183; Gaps 31;  
 QY 9 WGIALLALPPGAA---STOVCTGDMKRLPASEPETHLMDLRHLYOGGVVGNILETY 64  
 DB 8 WTVVSLVAAVGVQSDSQVAGTENKLSLSDEQYALRKYENECVVMGNILETS 67  
 QY 65 LPTNASLFLQDIOEVGVVLIANOVQVPLRIYRGVLPEDFNVALVLDNGDPLN 124  
 DB 68 IEHNDLSFLRSVRETVGVVALNQFRPLPENTRIIRGYLYEDRALALFLYRKDG 127  
 QY 125 NTPPYVAGSPGLREIQLRLTEILKGVLIQNPOLCYODTILMKDIFHKNOALTLI 184  
 DB 128 NF-----GLQELGLKNLLEILNGGVVYVQNNFLCYADTIHMDIYRNWPNLTLV 178  
 QY 185 DTRNSPACHPCSPMCKGSRGCESSDDCSLTRVYVAGGC-ARCKGPLETDCHEQCAAG 243  
 DB 179 STNGSSGCCRCHKSGCTG-RCMGPTENHCOTLTRVCAECCDCRCGPFVSDCHRECAAG 237  
 QY 244 CCGPKSDCLACLFHFGSGICELHCPALVTYTDTPESMPNDEGRVTFGASCVTACPFVY 303  
 DB 238 CGSPKDTDFACMKNFDSACVTCQPPOTVYVNPPTTFQLEHNNNAKTTGAFVYKACPFNF 297  
 QY 304 LSTDVSGCTLVPCPLHNOEYTABDGTORCEKSKPCARVYCYGMQYIKANSKFTIGTELE 363  
 DB 298 V-VDSSSCVRACPFSSIMEY-ENGKIMKPCPTDICPKACDGIIGTSLMSAQVDSNDIK 355  
 QY 364 PAGCKKIFGSLAFLESPFGDPAASNTAPLPOLQVFELEITGVYLSAMPDLSPLS 423  
 DB 356 FINTKIKINLFLVTGHHGDPYNAIEADPEKLVFVRIETITGLINQSPPVMTDPS 415  
 QY 424 VFQNTQVIRGLIHNGAVSLTQGLGISTNLGLRSURELSGIALITHNTHLCFVATVMD 483  
 DB 416 VFSNLVTIGRVLVYSLSLIKQGGITSLQGSLEISAGNIYITDNSNLGYHTIMWT 475  
 QY 484 QLFRRPHQALLTANRPDECVGEGLAGHQLCARHGCMGPGPTQCVCNSQPIRGCECVBE 543

Db 476 TLFSTINRIVRNDRKRAENCTAENGVNCHLSSDGCWGPBGDOCLSCRFRGRICIES 535  
 QY 544 CAVLOGLPREYVNAHCHLPCHEPCOP-GNGSVTCFGEPAADQCAVCAHYKDPFCYARCP 602  
 Db 536 CNLYOGEFFEFENGSTCYECPCQCKMEGGLTICGPGPDNCTKSHKRDGNCYCKEPPD 595  
 QY 603 GVKPDLISYMPKIPDEEGACQPCPINCTHSCVDDDKGC-----PAQRASPL 651  
 Db 596 GLOGANSF--IFKYADPDRECHCHPHNCTGCGNGCPTSHDCLYYPTWGTSTLFGHAR-TPL 652  
 QY 652 TSVSAVY-GILLVVLGVVPGILLKRRQOKIRKXTMRLLLOETLVEPLTPSGAMPQA 710  
 Db 653 --IAGVIGGLFVLVGLTFAVYRRKSIK-KKALRFL-ETLVEPLTPSGAMPQA 708  
 QY 711 QMRILKETELRKVKVLSGAGFPNNFTVSFWLRVP-----KVSASHLEVRNENTSPPKANK 765  
 Db 709 QRIKETELRKVKVLSGAGF---TYKKG:WVEGETVKIPVA-IKLNETTSPKANK 763  
 QY 766 ELIDEXYVWAGVSPYVSRLLGICLTSTVQLVTCMPYGCILDHRENRGRIGSODLLNW 825  
 Db 764 EFMDEALIVASWDHPLVLLGVCLSPITQLTQMLPGCLLEYVHEKNDTNGSOLLNW 823  
 QY 826 CMQIAKMSYLEDELVLRDLAARVLYKSPNHVKITDGLARLLIDETEHADGKVP 885  
 Db 824 CVOIAKGMVLEERLVRDLAARVLYKSPNHVKITDGLARLLIDETEHADGKVP 883  
 QY 886 IKMALESLRRRFTHOSCVMSYVTWELMTFGAKPYDGPAREIPDLLEKGERLP 945  
 Db 884 IKMALESLRRRFTHOSCVMSYVTWELMTFGAKPYDGPAREIPDLLEKGERLP 943  
 QY 946 ICTIDVYIMVCKMIDSECRPRELSEFSRMDRPPRFVIGNEO-LGASPLDSTF 1004  
 Db 944 ICTIDVYIMVCKMIDSECRPRELSEFSRMDRPPRFVIGNEO-LGASPLDSTF 1003  
 QY 1005 YKSLLDDMDGLVDAEYLYVQOGFCPPDPAAGAGVHHRSSSTRSGGDTLLGLE 1064  
 Db 1004 FQNLDEEDLEDMDMAEYLYV-QAFNIPP-----IYTSRARIDSKRS-----EIGH 1051  
 QY 1065 PEEERAPRS-----PLAP-SEAGSDVFDGDLGMMAA 1095  
 Db 1052 PPVATPWSGNQFYVRDGGFAAEQGVSPYRAPSTITPPAPVACATKTEIFDSCCNTL 1111  
 QY 1096 KGLQSLPTHPDPLORYSEDPVPLPS-----ETDGYVAPLTGSPQEVYVNDPVAPO 1148  
 Db 1112 KRPVAPVHGEDSSSTRYSADPTVFAPERSPRGELEDEGMTMRKPKQEVLYNPE---- 1167  
 QY 1149 PPSPREGLPAPRPPGATLERAKTILSPCKNGVYKDVFAFGAVENPEYLTPOGGAAPQH 1208  
 Db 1168 -----ENPFVSR-----KNGDLQ-----ALDNPEYHNASNG----- 1194  
 QY 1209 PPPA-----FSPAFLVYVMDQDPPRGA-- 1232  
 Db 1195 PPKADEVYNEPLVNTFANTTIGKAEYLNKNNILSPKAKKAFDNPDPWNSHSLPRSTLQ 1254  
 QY 1233 PPSTFKGTF-----AENPEYL 1249  
 Db 1255 HEDVLOEYSTKYFYKONGRIRPIVAENPEYL 1285  
 RESULT 7  
 ERB4\_RAT STANDARD; PRT; 1308 AA.  
 ID ERB4\_RAT  
 AC Q62956; Q922N7; (Rel. 37, Created)  
 DT 15-DEC-1998 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Receptor protein-tyrosine kinase erbB-4 precursor (EC 2.7.1.112).  
 GN ERB4 OR TYRO-2.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_Taxid=10116;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC TISSUE=Heart;  
 RX MEDLINE=98221155; PubMed=9553078;  
 RA Zhao Y.-Y., Sawyer D.R., Baliga R.R., Opel D.J., Han X.,  
 RA Marchionni M.A., Kelly R.A.,  
 RT "Neuregulin promotes survival and growth of cardiac myocytes.  
 RT Persistence of Erb2 and Erb4 expression in neonatal and adult  
 RT ventricular myocytes."  
 RL J. Biol. Chem. 273:10261-10269(1998).  
 RN [2]  
 RP SEQUENCE OF 848-901 FROM N.A.  
 RC TISSUE=Sciatic nerve;  
 RX MEDLINE=91222560; PubMed=2025425;  
 RA Iai C., Lemke G.,  
 RT "An extended family of protein-tyrosine kinase genes differentially  
 RL expressed in the vertebrate nervous system";  
 RN Neuron 6:691-704(1991).  
 RP SEQUENCE OF 1031-1198 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Spinal cord;  
 RX MEDLINE=97184212; PubMed=9030624;  
 RA Carroll S.L., Miller M.L., Frohert P.W., Kim S.S., Corbett J.A.,  
 RT "Expression of neuregulin and their putative receptors, Erb2 and  
 Erb3, is induced during Wallerian degeneration."  
 RL J. Neurosci. 17:1642-1659(1997).  
 CC -1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULINS, NRG-  
 CC 2, NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, RETACELULIN AND  
 CC NNAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION.  
 CC NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN (BY SIMILARITY).  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERB  
 CC RECEPTORS (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- TISSUE SPECIFICITY: PREFERENTIALLY EXPRESSED IN THE DEVELOPING  
 CC NERVOUS SYSTEM. EXHIBITS DISTINCT AND HIGHLY REGIONALIZED PATTERNS  
 CC OF EXPRESSION IN THE ADULT BRAIN, WHERE IT IS MAINLY FOUND IN THE  
 CC RETICULAR NUCLEUS OF THE THALAMUS. VERY LOW LEVELS IN KIDNEY, ADI  
 CC HEART.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE  
 CC RESIDUES (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb.sib.ch](mailto:license@isb.sib.ch)).  
 CC -----  
 DR EMBL; AF041838; AAD08899.1; -  
 DR EMBL; U52531; AAC53051.1; -  
 DR HSSP; P11362; IFGK.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR InterPro; IPR004019; VLP\_motif.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; pkinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR Pfam; PF02757; VLP\_2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR PRODOM; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00261; Fv; 4.  
 DR SMART; SM00219; TYKc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 KW Transmembrane, Glycoprotein, Multigene family, Receptor, Signal,  
 KW Transferase, Tyrosine-protein kinase, ATP-binding, Phosphorylation,  
 FT SIGNAL  
 1  
 25  
 POTENTIAL.



FT CHAIN 26 1308 RECEPTOR PROTEIN-TYROSINE KINASE ERBB-4.  
 FT DOMAIN 26 651 EXTRACELLULAR (POTENTIAL).  
 FT TRANSMEM 652 675 POTENTIAL.  
 FT DOMAIN 676 1308 CYTOSOLASTIC (POTENTIAL).  
 FT DOMAIN 186 334 CYS-RICH.  
 FT DOMAIN 496 633 CYS-RICH.  
 FT DOMAIN 718 985 PROTEIN KINASE.  
 FT NE BIND 724 732 ATP (BY SIMILARITY).  
 FT BINDING 751 751 ATP (BY SIMILARITY).  
 FT ACT\_SITE 643 643 BY SIMILARITY.  
 FT DISULFID 189 197 BY SIMILARITY.  
 FT DISULFID 193 205 BY SIMILARITY.  
 FT DISULFID 213 221 BY SIMILARITY.  
 FT DISULFID 217 229 BY SIMILARITY.  
 FT DISULFID 230 238 BY SIMILARITY.  
 FT DISULFID 234 246 BY SIMILARITY.  
 FT DISULFID 249 258 BY SIMILARITY.  
 FT DISULFID 262 269 BY SIMILARITY.  
 FT DISULFID 293 304 BY SIMILARITY.  
 FT DISULFID 308 323 BY SIMILARITY.  
 FT DISULFID 326 330 BY SIMILARITY.  
 FT DISULFID 503 512 BY SIMILARITY.  
 FT DISULFID 507 520 BY SIMILARITY.  
 FT DISULFID 523 532 BY SIMILARITY.  
 FT DISULFID 536 552 BY SIMILARITY.  
 FT DISULFID 555 569 BY SIMILARITY.  
 FT DISULFID 559 577 BY SIMILARITY.  
 FT DISULFID 580 589 BY SIMILARITY.  
 FT DISULFID 593 614 BY SIMILARITY.  
 FT DISULFID 617 625 BY SIMILARITY.  
 FT DISULFID 621 633 BY SIMILARITY.  
 FT MOD\_RES 1162 1162 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1188 1188 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1258 1258 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT MOD\_RES 1284 1284 PHOSPHORYLATION (AUTO-) (BY SIMILARITY).  
 FT CARBOHYD 138 138 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 174 174 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 253 253 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 358 358 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 410 410 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 473 473 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 495 495 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 548 548 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 576 576 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 620 620 N-LINKED (GLCNAC) (POTENTIAL).  
 FT CARBOHYD 1062 1062 S -> N (IN REF. 3).  
 FT CONFLICT 1080 1082 PYT -> SYR (IN REF. 3).  
 SQ SEQUENCE 1308 AA; 146957 MW; D944B0996A08B41 CRC64;

Query Match 41.64; Score 2833; DB 1; Length 1308;  
 Best Local Similarity 43.94; Pred. No. 2e-144;  
 Matches 594; Conservative 194; Mismatches 394; Indels 170; Gaps 31;

QY 1 MELA-ALCRWGLL--ALLPFGAATQVCTGTDMLRLPASPEHLMDRLHYQCGVQ 57  
 Db 1 MKLATGLMVGSLVAVARTVQPSASQSVCACTENKLSLSLDEQYRALRYKYECEVVM 60  
 QY 58 GNLEIYLPFNASTSLFODIOEVOGYVLLAHNOVRVQRLRIYVGTQLFENYLAFL 117  
 Db 61 GNLEITSTIEHNRPLSRISREVTGYVALNQFRPLFENRITIGTLYEDRYALALF 120  
 QY 118 DNGDPLNNTTPVTGASFGRLRELQSLTEILKGVLIQRPOLCQDTILMKDFHKN 177  
 Db 121 LNVKXQGNF-----GLQELGLNKLTEILNGVYVDQNKFLCAADITHQDIYRNM 171  
 QY 178 QALALTLIDTKRSRACHPCSPMGSKGCKWGBSSSDQSLRTTVAGGC-ARCKGPLETDC 236  
 Db 172 PSMWTLVSTIGSSGCGRCKSKCTG-RCWGPTENHCQTLRTVACEDCDRCGVPYVSDC 230  
 QY 237 HEOCAAGCTGPKSDCLACJLHFNHSGICELHCPALVTYNTDTFESPNPEGRYTFGASCV 296  
 Db 231 HRCAGGCGSPKDTDFACACNPFNDGACVTCQCPOTVNPFTTQLEHNNAKYTYGAFV 290

QY 297 TACPYNLTSTVSGCTLVCPILHNOEYATADGTORCEKSKPCARVCGYLGMOYIKANSKF 356  
 Db 291 KCCPKNFV-VDSSCCVACAPSSKMEV-EENIKKCKCTDLCRACGIGIGSLMSQTV 348  
 QY 357 IGITELEPAGCKKI FGSIALPLPESFPDDPASNTAPLOPELOVETLEITGYIYISAMP 416  
 Db 349 DSNIDKRFNCKINGNLIPLVGTIGHGPYNAIDAIPEKLVNRTVREITGFINICTWP 408  
 QY 417 DSLPDLVSFQNLVIRGILHNGAVSLTLQGLGSIWGLRSLRGLSGSLAIHHNTHCF 476  
 Db 409 PNMDFSVFSLVLTGGKVLVSGSLTLKQGLITSLOPQSLKESAGNIYITNSNLCY 468  
 QY 477 VHTVPMQDLFRNPHQALHTANRPEDECVGEGLAGHOLCARGHMGCPPTCVANCQGLR 536  
 Db 469 YHTINMTLFTSTVQRIYIRNBRANENCTAGWCMHLCSDGCMGPPDCLSCRERS 528  
 QY 537 GQECYEBEVRVLOGLPREVYNARHCLPCHPECP-ONGSVTCFGEADQVACAYKDPF 595  
 Db 529 GKICIESCNLYGDEFERENGISICVECDQCEKXEDGLTGHGPPDCTKSHFKQGN 588  
 QY 596 CVARCPGVRKPDLSYMPIMKPPDEGACQPCPINCTSHCVLDLQKC-----PA 644  
 Db 589 CVEKCPDVLQGANF--IFKADQDRCHPCHPCTQCCNGPSTHDCIYPTWGHSTLP 646  
 QY 645 EORASPLTSYSAV-GILLVVLGVVFGILIKRQOKIRRYMRLLQETELVEPLTS 703  
 Db 647 HAR-TPL--IAAGVIGLFIIVIMALTFAYVVRKSIK-KRALRRFL-ETELVEPLTS 701  
 QY 704 GAMPQAOMLRKTELRLKVKVVLGSGAFGNFVSPFLRVP-----KVASHLEVREN 758  
 Db 702 GTAPQAOURLRKTELRLKVKVVLGSGAFG---TVYGIWVPEGETYKIPVA-IKILMET 756  
 QY 759 TSPKANKEILDEAYVMAGVGSPPYVRLIGICTSTVQVLTQMLMEYCYLDHVRNRLG 818  
 Db 757 TGPANVEMFDEALIMASVDRPHLVRLGVCSTPTQVLTQMLHGGCLEVHEHKNIG 816  
 QY 819 SQDLINMCQIAKMSYIEDVLRHDLAANVLYKSPNNHYKITDFGLARLLDDETEYH 878  
 Db 817 SQLINMCVQIAKMMYLEERLVHRDLAANVLYKSPNNHYKITDFGLARLLDDETEYH 876  
 QY 879 ADGGVPIKMMALLESILRRFTHSDVSVYVWELMTFGAKPYDGIPIAREIPDLLEKG 938  
 Db 877 ADGGMPFKMMALBCHIRKFTHSDVSVYVWELMTFGAKPYDGIPIAREIPDLLEKG 936  
 QY 939 ERLPPICTIDVYIMVYKMMIDSECRPRELIVSEFSRYARDPQRFVITQND-LGPA 997  
 Db 937 ERLPPICTIDVYIMVYKMMIDSECRPRELIVSEFSRYARDPQRFVITQND-LGPA 996  
 QY 998 SPLDSTFYRSLLEDMDMDLVDADAEYLVPOGFCFCDP----- 1035  
 Db 997 SPNDKPFQNLDEEDLDDMDADAEYLV-CAFMIPPEIYTSRRIRDSNRSEIGHSPPA 1055  
 QY 1036 APGAGVHHRHRSSTRSQGDULTGLEPSEEBAPRSLAPSEGAGSDVFDGLMGCA 1094  
 Db 1056 YTPMSGQFVYQDGGFATQGC--MPMYTATSTIPAPVA-QGATAEMFDDSCNGT 1110  
 QY 1095 AKGLQSLFTHPSPLOKRSDEPTVPLRS-----ETDGYAPLTCSPQRYVQDVPR 1147  
 Db 1111 LRKPVPHVQDSSSTORSADPTVAFAPRNRAELDEGYMTPMHDKREKQYLVNVE--- 1167  
 QY 1148 QPSPREGRPLPAASPAGATLERAKTSLSGKGVVVDYAFAGVAJENFEYLLPQGAAPQ 1207  
 Db 1168 -----EHPFVSR-----KNGDLQ-----ALDNEFEYHASSG----- 1194  
 QY 1208 HPPPA-----ESPADFNIYWDQDPPEGA- 1232  
 Db 1195 -PPKAEDEVNPEPLNTFTNALGNAEYMKSLSLVEKAKKADNDNDYHSLPPESTL 1253  
 QY 1233 -PPSTFKGTP-----AENPEYL 1249  
 Db 1254 QHPDYLQDYSTKYFYKONGRLRPIVAENPEYL 1285

RESULT 8  
XMRK\_XIPMA STANDARD: PRT: 1167 AA.  
ID XMRK\_XIPMA PRT: 1167 AA.  
AC P13388;  
DT 01-JAN-1990 (Rel. 13, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Melanoma receptor protein-tyrosine kinase precursor (EC 2.7.1.112).  
GN XMRK OR TY.  
OS Xiphophorus maculatus (Southern platyfish).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
OC Acanthomorphi; Acanthopterygii; Percomorpha; Atherinomorpha;  
OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
OK NCBI\_TaxID=8083;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90015140; PubMed=2797166;  
RA Witbrodt J., Adam D., Maltschek B., Maueier W., Raulf F.,  
RA Telling A., Robertson S.M., Scharlt M.;  
RT "Novel putative receptor tyrosine kinase encoded by the melanoma-  
inducing tu locus in Xiphophorus.";  
RL Nature 341:415-421 (1989).  
RN [2]  
RP REVISION TO 515.  
RA Scharlt M.;  
RT Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.  
CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
CC tyrosine phosphate.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.  
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC -----  
DR EMBL; X16891; CAA34770.2; -.  
DR PIR; S06142; S06142.  
DR HSP; P11362; 1FGK.  
DR InterPro: IPR000494; EGFR\_L domain.  
DR InterPro: IPR000719; Euk\_Dkinase.  
DR InterPro: IPR002174; Eukin-like.  
DR InterPro: IPR002290; Ser\_Chr\_Pkinase.  
DR InterPro: IPR001245; Tyr\_Pkinase.  
DR Pfam; PF00069; Pkinase; 1.  
DR Pfam; PF00757; Furin-like; 1.  
DR Pfam; PF01030; Recep\_L domain; 2.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Euk\_Pkinase; 1.  
DR SMART; SM00261; FU; 5.  
DR SMART; SM00220; S\_TKC; 1.  
DR SMART; SM00219; TYKIC; 1.  
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
DR PROSITE; PS50011; PROTEIN KINASE DOM; 1.  
KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;  
KM Tyrosine-protein Kinase; ATP-binding; Phosphorylation; Proto-oncogene.  
FT SIGNAL 1 25  
FT CHAIN 26 1167  
FT DOMAIN 26 642  
FT TRANSMEM 643 665  
FT DOMAIN 666 1167  
FT DOMAIN 710 977  
FT NP\_BIND 716 724  
FT BINDING 743 743  
FT ACT\_SITE 835 835  
BY SIMILARITY.

FT DISULFID 195 204 BY SIMILARITY.  
FT DISULFID 199 212 BY SIMILARITY.  
FT DISULFID 220 228 BY SIMILARITY.  
FT DISULFID 224 236 BY SIMILARITY.  
FT DISULFID 237 245 BY SIMILARITY.  
FT DISULFID 241 253 BY SIMILARITY.  
FT DISULFID 256 265 BY SIMILARITY.  
FT DISULFID 269 296 BY SIMILARITY.  
FT DISULFID 300 311 BY SIMILARITY.  
FT DISULFID 315 330 BY SIMILARITY.  
FT DISULFID 333 337 BY SIMILARITY.  
FT DISULFID 304 513 BY SIMILARITY.  
FT DISULFID 508 521 BY SIMILARITY.  
FT DISULFID 524 533 BY SIMILARITY.  
FT DISULFID 537 553 BY SIMILARITY.  
FT DISULFID 556 569 BY SIMILARITY.  
FT DISULFID 560 577 BY SIMILARITY.  
FT DISULFID 593 615 BY SIMILARITY.  
FT DISULFID 618 626 BY SIMILARITY.  
FT DISULFID 622 634 BY SIMILARITY.  
FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 201 201 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 398 398 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 417 417 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 501 501 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 576 576 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 621 621 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 1167 AA; 129934 MW; 479380749DCID55A CRC64;  
Query Match 37.8%; Score 2575.5; DB 1; Length 1167;  
Best Local Similarity 44.0%; Pred. No. 1.1e-130;  
Matches 561; Conservative 168; Mismatches 392; Indels 155; Gaps 33;  
QY 4 AALCRGALLALLPPEAST---QYCTDMKRLRPAPEHLDMLRLYGGCCYVQGN 59  
DB 8 AALLQ--LLVLSISRCSTDPDRKVCQSTSNQMTN--LDNHYLMKMKYSGCNVLEN 62  
QY 60 LELTYLPINASLFLDIDIEVQGYVLIANQVQVPLQRLRVGTQLPFDNYVALVDN 119  
DB 63 LEITYQENODLSFLQSIQEVGYVLIANNEVSTIPLVRLIRGQNLVEGNFTLLWMN 122  
QY 120 GPELNTTPVTASRGLRELQRLSTETLTKGVLIQRNPOLCYQDTILMKOIFRNNOL 179  
DB 123 YOK-NPSSP--DYVQGLKQLSLNLTETLSGVKVSHPPLCNVETIMMDIVXTSNP 179  
QY 180 AALLIDPNSRACHPSPSPKSGSRGSESEDCQSLTRTVCAAGC-ARCKGLPPTCCHE 238  
DB 180 TNNLIPHAFERQCKDHCQCVGSCWAPRGHCQKFTKLCAEQCNRRRGKRPIDCNE 239  
QY 239 QCAAGCTGKASDCLALHFNHSGICELCPALVYNTDTFESMPNBEGRYFGASCYTA 298  
DB 240 HCAGGCTGPRATDCIACRDENDGCTCKDPPKXIYDIYSHQVNVNPNKIYFGAACVKE 299  
QY 299 CPYNYLSTDVSGCTLVCP-CHNQEVTAEDTQCEKSKRCACVCGIGMQYIKANSKFIG 358  
DB 300 CERNYVIE-GACVASCAGMLEVD-ENKSKCKCDGVCPVCGIGGSI-SYTIANV 356  
QY 359 ITEL-EPAGCKKIFGSLFLPESFDGDPASNTAPLQPEQLQYFETLEETLGYLYSAMPD 417  
DB 357 STNINSFSNCTKINGDILNRSNFGDPDPRHYKIGTMDPEHMLNLTIVKEITGLVIMWPE 416  
QY 418 SLPLDSYQNLQVINGRILHNAYS-LTLQIGISWLGIRSRBELSGALIHNTHLCE 476  
DB 417 NNTSLSVFNNLEIIGRTTFSGFSFVVQVYVHQLGLRSIKREVSAGVAILKNTLQRY 476  
QY 477 VHTVPMDQLFRNPQALLHTANRPEDCEVGEGLACHQCARHGCVGPGTQVNCSEFLR 536  
DB 477 ANTIMMRIRFREDSISYDART-----ENQTCNMBCSEBGCNGPBTWKVCSGLAHDR 529  
QY 537 GGECEVBECEVTLQGLFREYVNAHCLPCHPBCQPNQNGSVTCFGEPAEDQVACAHYKDPFPC 596

```

Db 530 GGRCVASCNLLQGEPRQAVDGRVQCHQCELVQDLSLTCVGPBPANCSKSAHPQDQOC 589
Qy 557 VARGSGVPRDLSTYPIKMFPEBAGACOPCPINCHSCVDDDKCCPEAGASPLTISVS 656
Db 590 IPRCPGILGDDTL-IMKADKWCQCCOPCHONCQSGPGLSGCRD-IVSHSLAVG 647
Qy 657 AVVGLLVVGLVGVGILIKRQOKIRKYMRLQETELVEPLTPSAMPQAMRLIK 716
Db 648 LVSGLLITVALLVILVLRRLIK-RKRTICLLQKELEVEPLTPSQAQNAQAFRLIK 706
Qy 717 ETELKRVKVLGSGAGCFNNFTV--SPW-----LAVPKVSAHLEVLRENTSPKANKIL 768
Db 707 ETEFKKRVLGSGAGG---TYXKGLMPDGENIRIPVA---IVLEAETS PKVNOEVL 758
Qy 769 DEAYVAGVSPYVRLIGICTSTYVOLTOLMPGCLLDHYRENRGLSGSDLLNMCQ 828
Db 759 DEAYVMSVDHVCRLQICLTSVQVLTOLMPGCLLDYRQOERICQOMLNCVQ 818
Qy 829 IAKGMSYLEDVLRHDLAARNVLYKSPNHVKITDFGLARLLDIDETBYHADGKVPRIK 888
Db 819 IAKGMVYLEERLVRHDLAARNVLLKNPNHVKITDFGLSKLTADKEKYQADGKVPRIK 878
Qy 889 MALESILRRRTFQSDWSYGVTWELTFGAKPYDGI PAREIDPLEKGRLOPPIC 948
Db 879 MALESILQWTTYHSDWSYGVTWELTFGSKPYDGI PAKEIASVLENGERLPPIC 938
Qy 949 IDVYIMVYKQWIDECAPREFELVSEFSRMADQREFVINCEDLGPASPLDSTFYSL 1008
Db 939 IEVYMIKQWIDESSRPFREFELVSEFSQMARDSRYLVIG--NLPSLDRRLSRL 995
Qy 1009 LEDDMGDLVDAEELVLPQGGFCPCDPAPAGAGVHHRRSSSTSGGDLTLGLEPSEE 1068
Db 996 LSSDD--DVADDEVLLPYKRI-----NFGGS----- 1020
Qy 1069 EAPRSPLAPSEAGSDVDFDGLMGANGLOSLPHTDPSPLQRYSEDTV-PLPSETGY 1127
Db 1021 --EPCLPPTGH-----PVENSIILRNISDTQNALEKDLGH 1056
Qy 1128 VAPLTCSQPEYVNPDPVRPP--PSPRE-----GRLP-AARPAGATLERAKTUS 1174
Db 1057 -----EYVNPQGETSRLSDIYNPNVEDLTDGMPVSLSSQEAETNFRPELTIN 1106
Qy 1175 FGRKGVKDVFAFGAVENPEYLTPOGAAPQHPHPPAFADMLYVDDCPBERGAP 1234
Db 1107 TNQNSL---PLVSSGSMDDPDV---QAG-----YQAAF-----LPQTGALT 1141
Qy 1235 STFKGTPTAENPEYIG 1250
Db 1142 GNGMFLPAENLEYIG 1157

RESULT 9
ERBB3_HUMAN STANDARD; PRT; 1342 AA.
ID ERBB3_HUMAN
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3) (Tyrosine kinase-type cell surface receptor HER3).
GN ERBB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=90083234; PubMed=2687875;
RA Kraus M.H., Issing W., Miki T., Pinescu N.C., Aaronson S.A.;
RT "Isolation and characterization of ERBB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors."

```

```

FT BINDING 742 742 ATP (BY SIMILARITY)
FT ACT_SITE 834 834 BY SIMILARITY
FT DISULFID 186 194 BY SIMILARITY
FT DISULFID 190 202 BY SIMILARITY
FT DISULFID 210 218 BY SIMILARITY
FT DISULFID 224 226 BY SIMILARITY
FT DISULFID 227 235 BY SIMILARITY
FT DISULFID 231 243 BY SIMILARITY
FT DISULFID 246 255 BY SIMILARITY
FT DISULFID 259 286 BY SIMILARITY
FT DISULFID 290 301 BY SIMILARITY
FT DISULFID 305 320 BY SIMILARITY
FT DISULFID 323 327 BY SIMILARITY
FT DISULFID 500 509 BY SIMILARITY
FT DISULFID 504 517 BY SIMILARITY
FT DISULFID 520 529 BY SIMILARITY
FT DISULFID 533 549 BY SIMILARITY
FT DISULFID 552 565 BY SIMILARITY
FT DISULFID 556 573 BY SIMILARITY
FT DISULFID 576 585 BY SIMILARITY
FT DISULFID 589 610 BY SIMILARITY
FT DISULFID 613 621 BY SIMILARITY
FT DISULFID 617 629 BY SIMILARITY
FT CARBOHYD 126 126 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 250 250 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 353 353 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 408 408 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 414 414 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 437 437 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 469 469 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 522 522 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 566 566 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT CARBOHYD 616 616 N-LINKED (GLCNAC . . .) (POTENTIAL)
FT VARSPPLIC 141 183 ELSGGVYIEKNDKCHMDTDMDIYVDRABEYVQDNGR
SC -> GQEPVPSGLTPQADWTLDDDEPLLTLSASSK
VPTLAIV (IN SHORT ISOFORM)
FT VARSPPLIC 184 1342 MISSING (IN REF. 2)
FT CONFLICT 560 560 E -> G (IN REF. 2)
FT CONFLICT 1064 1064 E -> G (IN REF. 2)
SQ SEQUENCE 1342 AA, 148097 MW, 7201E7F66C4374BD CRC64;

```

```

Query Match 34.0%; Score 2316.5; DB 1; Length 1342;
Best Local Similarity 39.5%; Pred. No. 9.5e-117;
Matches 520; Conservative 196; Mismatches 459; Indels 143; Gaps 36;

10 GLLTALLPAGAA--STGYCTGTDMKRLPASPETHLDMLRLHYGCGVVOGNLELTLYPT 67
11 GLPFLARGSEVGNQAVCPGTGLNGSTGDAENQYOTLYKLYCEVVMGNLELTVLTHG 70
68 NASISFLQDIOEVQGYVLIANQVQVEFLQRLIRVGTQLFEDNYVALAVLDNGDPLNNTT 127
71 NADLSFLQWIREVTGYVLVAMNEFSTLPLPMLRVYRGQVYDGKPAIFVM-----LNYNT 125
128 PVTGASPGGLRELQRLSTELLKGVLIQGNPQCYOTILMKDIFHKNQDLATLTDN 187
126 ----NSSHALRQRLTQLTLLISGVYTEKNDKLCMDITMDRIIVDRD---AEIYVKD 178
188 RSRACHPCSPMKSGRCWGESSEDCSLTRTVCAAGC-ARCKGFLPTDCHEQCAAGCTG 246
179 NGRSCPCHVEYCKG-RCMGPGSEDCQTLTITICAPQCNHGFCFNGPNQCCDEACGCGSG 237
247 PKHSCTLACLHPNHSIGTELHCPALVYNTDTFESMENPEGRYFPGASCYTACPNYLTST 306
238 PQDDTDFACRFHFNNSGACVPRCPQLVYVYKTLFQLEPFPATKRYGVGCVAVSCPHNFV-V 296
307 DVSGCTIVCPPLHNOEVAEDGNORCEKSKPCAVCGVLGMQYIKANSKF--IGITILE- 363
297 DQTCVTRACFPDKNHEVD-KNGIKNCEPFGGICPKACBETG-----SGSRFTVSSNIDG 350
364 FAGCKKIFGSLAFLPESEFDGDPASNTAPLQPEQLOVEFTLEITGYLYISAMPDLPDLS 423
351 FVNCTKILGNLDFLTGLNGDPMHKIPALDPEKLNVPFTVAIEITGYLINTIGMWPMPHMFNS 410

```

```

QY 424 VFQNLQVIRGRILHNGAYS-LTLQGLISMLGLRSIRELGSGLALLHHNTHLCFVHTVPM 482
DB 411 VFSNLTLTIGRSLYNRFPFSLIMKNLNTVTLSTGRSLKEISAGRIYISANQLCYHSHLW 470
QY 483 DQLEFRNHQALLHTA-NRPEDECVGSELAHQICAGHGMWGPPTQCVNCSQLRQGEV 541
DB 471 TKVLRGTEERLDIKHNRPRDVCBSKVCDDPLCSSGCGPQGGQCLSRANSRGGVCV 530
QY 542 EECRVLOGLPREYVNAHCLPCHPECQPNQSGYTCFPEADQCVACAHYVDPFCVACRP 601
DB 531 THCNFLNGEPRFAHEAECSCHPEQPMGTATCNQSSGSDTCAQCHFFDGRHVCSSCP 590
QY 602 SGVPRDLSYPIKFPPEBACQPCPINCASHCVDDLDDKCPABQRA---SPLTSVSA 657
DB 591 HGVLG-AKSPRYKYPVQNECRPCHENCTQGGCKGELDQCLQTLVLJCKTHLTVALTV 648
QY 658 VGLGLLVVGVVFGILIKRROKIR-KYMRRLDQTELVPEPLTPSGAMPNOQWILK 716
DB 649 IAG--LVYIPMLGCTLYRGRRIQKRAMRYLERGESIEFLDPS-EKANKYLARIK 705
QY 717 ETELKRYKVLGSAFGFNNFTVSFWLRVPKVSASH---LEVLRENTSPRANKELLIDEAY 772
DB 706 ETELKRYKVLGSAFGF---TVHKGVMIPGESIKIPVCIKVIEDKSGRQSPQAVTDHML 761
QY 773 VMAGVSPYVSRILGICLSTVQVNTQMLPYGCLLHVRENRGRISQDILLNMCQIATK 832
DB 762 AIGSLDHAHIVRLGLCPGSSILQVLYQPLGSLDHRQALGPQLLLNNGVOIATK 821
QY 833 MSYLEDVRLVHRDLAARVNVKSPNHKIDFGLARLIDETEVHADGKVPKIMALE 892
DB 822 MYLIEBGMHRLAARVNVKSPQOVAVDLPVADLLPPDDQQLYSEAKTIKIMALE 881
QY 893 SIIRRRTHOSDVSVCVTVWELMTFGAKPYDGI PARIEPLDLEKGRLLPQPICTIDVY 952
DB 882 SIHFQKXTHOSDVSVCVTVWELMTFGAEVYAGRLAEVADLLEKGRLLAQPOICTIDVY 941
QY 953 MIMVKCMIDSECRPRELVSFSPMARPPQCFVYIOWEDLQPA---SPLDSTFYSLL 1009
DB 942 MIMVKCMIDENIRPTFKELAMFTTMADPPRYLVYIKES-GRGLAPGEPPGLTNKKL 1000
QY 1010 EDDDMGDLVDAEEYLVPOQGFPCPDPAAGAGVHHRHRSSTRSGGDLTLLGP-SEE 1068
DB 1001 EEVELBEPLDLDLDEAED-----NLATTTIGSALSLVGTILNR 1041
QY 1069 EAPRSPPLABEGAGSDVFDGDLGMGAAGLQSLPTPD-ESPLORYSEDPVPLP----- 1121
DB 1042 RGSQSLSPSSGY-MPMNQNLGESCQESAVSSSERCPVSLH-----PMPROGLAS 1094
QY 1122 SETDGUYA-----PLTCSPOPE-----YVNOQPVPRQPPSPREGP----- 1156
DB 1095 ESSEGHVYTGSEALQKVSMSCRSRSSRSPRPGDSAYHSQKRSLLTPVTPPLSPGLLEE 1154
QY 1157 -----LPAARPGATLERAKTLP-S-GKNQV-----KDVAFAGAVENPEYLVPOGGA 1203
DB 1155 DVNGVYVPTDHLKGTSSREGTSLVGLSVLTGTEDEED-----EEVEYMNRRRH 1206
QY 1204 APOPHPEPAFSPAFDLVYWD-----QDPPEKGAPESTFKGPTIANPEYL 1249
DB 1207 SP-PHPPRESSLELGEIYMDVGSDLASLGSSTQSCPILHPVIMPTAGTTPDDEYEM 1263

```

```

RESULT 10
ERB3 RAT STANDARD; PRT; 1339 AA.
AC 062759; 062955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Receptor protein-tyrosine kinase erbB-3 precursor (EC 2.7.1.112)
DE (c-erbB3).
GN ERB3
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;  
 RX MEDLINE=96096535; PubMed=8522190;  
 RA "Heller N.J., Kim H.-H., Greaves C.H., Sierke S.L., Koland J.G.;  
 RT "Cloning of the rat ErbB3 cDNA and characterization of the  
 RL recombinant protein.";  
 RN Gene 165:279-284(1995).  
 RP REVISIONS TO 85, 513 AND 565.  
 RA Heller N.J., Koland J.G.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE OF 922-1097 FROM N.A.  
 RC STRAIN=Sprague-Dawley; TISSUE=Sciatic nerve;  
 RX MEDLINE=97184212; PubMed=9006624;  
 RA Carroll S.L., Miller M.L., Fromert P.W., Kim S.S., Corbett J.A.;  
 RT "Expression of neuroligin and their putative receptors, ErbB2 and  
 RT ErbB3, is induced during Wallerian degeneration.";  
 RL J. Neurosci. 17:1642-1659(1997).  
 CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NTRK  
 CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein  
 CC tyrosine phosphate.  
 CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS  
 CC (POTENTIAL).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE  
 CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.  
 CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES  
 CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF  
 CC PHOSPHATIDYLINOSITOL 3-KINASE.  
 CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.  
 CC This Swiss-Prot entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: U29339; AAC28498.2;  
 DR EMBL: U52530; AAC53050.1;  
 DR HSRP: P11362; IFCX.  
 DR InterPro: IPR000494; EGFR\_L\_domain.  
 DR InterPro: IPR000719; Euk\_pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF000657; pkinase; 1.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF01030; Recep\_L\_domain; 2.  
 DR PRINTS: PR00109; TYRKINASE.  
 DR ProDom: PDO00001; Euk\_pkinase; 1.  
 DR SMART: SM00261; FU; 5.  
 DR SMART: SM00219; TYK; 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; FALSE\_NEG.  
 DR PROSITE: PS00109; PROTEIN KINASE\_TYR; 1.  
 DR PROSITE: PS00011; PROTEIN KINASE\_DOM; 1.  
 DR Transmembrane; Glycoprotein; Multi-gene family; Receptor; Signal;  
 KW Transmembrane; Glycoprotein; Multi-gene family; Receptor; Signal;  
 FT SIGNAL 1  
 FT CHAIN 20 1339  
 FT DOMAIN 20 643  
 FT TRAMEM 644 662  
 FT DOMAIN 663 1339  
 FT DOMAIN 183 259  
 FT DOMAIN 707 964  
 FT NP\_BIND 713 721  
 FT BINDING 740 740  
 FT ACT\_SITE 832 832  
 FT DISULFID 186 194  
 FT BY SIMILARITY.  
 FT BY SIMILARITY.

FT DISULFID 190 202 BY SIMILARITY.  
 FT DISULFID 210 218 BY SIMILARITY.  
 FT DISULFID 214 226 BY SIMILARITY.  
 FT DISULFID 227 235 BY SIMILARITY.  
 FT DISULFID 231 243 BY SIMILARITY.  
 FT DISULFID 246 255 BY SIMILARITY.  
 FT DISULFID 259 286 BY SIMILARITY.  
 FT DISULFID 290 301 BY SIMILARITY.  
 FT DISULFID 305 320 BY SIMILARITY.  
 FT DISULFID 323 327 BY SIMILARITY.  
 FT DISULFID 500 509 BY SIMILARITY.  
 FT DISULFID 504 517 BY SIMILARITY.  
 FT DISULFID 520 529 BY SIMILARITY.  
 FT DISULFID 533 549 BY SIMILARITY.  
 FT DISULFID 556 573 BY SIMILARITY.  
 FT DISULFID 576 585 BY SIMILARITY.  
 FT DISULFID 589 610 BY SIMILARITY.  
 FT DISULFID 613 621 BY SIMILARITY.  
 FT DISULFID 617 629 BY SIMILARITY.  
 FT CARBOHYD 126 126 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 250 250 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 353 353 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 408 408 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 414 414 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 437 437 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 469 469 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 522 522 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 566 566 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CARBOHYD 616 616 N-LINKED (GLCNAC). (POTENTIAL).  
 FT CONFLICT 1028 1028 L -> P (IN REF. 3).  
 SQ SEQUENCE 1339 AA; 147545 MW; 0AA5F2402BBDFDIE CRC64;  
 Query Match 32.9%; Score 2244.5; DB 1; Length 1339;  
 Best Local Similarity 39.7%; Pred. No. 6.8e-113;  
 Matches 512; Conservative 173; Mismatches 437; Indels 167; Gaps 37;  
 3 LAALCRGGLLALLPQAA---STQVCTGDMKLRPASETHIDMKRLHYQCGQVQGN 59  
 7 LQVLC---FLSLARGSENGNSQAVCPGLINGSLVTDADNOYQTLKYLKCEVVMGN 62  
 60 LELVLPNLSLSDIGCEVGVVIAHNOVQVPLQRIYRGQLFEDNVALAVLDN 119  
 63 LEIVLTGNADSLFLQIREVTGYLVANNEFSVLPENLRVVGTVYQKRAIFVM-- 120  
 120 GDPPLNTTPYTGASPGGLRELOSLTEILKGVLIQRPOLCYODTILMKDIFHKNNOL 179  
 121 ---LNYNT---NSSHLRQLQKFTQLEILISGVYIEKNDKLCIMDTIDMRDIVRV-- 170  
 180 ALTLIDNRSRACHPCSPMKCKSRGCESEDDQSLTRYTCAGGC-ARCKGRLPTDCCH 238  
 171 GAETIVKRNNGANCPCHCEVCKG-RCWPGPDDQILTKTICAQCKNGRCGPPNQCCHD 229  
 239 QCAAGCTGPRGSDCLAFHNHSGICELHCPALVYNTDFFESMPNPEGRTYGASCVTA 298  
 230 ECAGGCGSGPDQTCFACRRRNDGACVCPREPLVYNKTLFOLPEPHHTYQYGVCVAS 289  
 299 CPNRYLSTDVGSCCTLVCPFLNNGVTLAEDGTQRECKSKPCARVYGL--GMQYIKANSKF 356  
 290 CPNHFV-VDOTFCVRACPDKMEVD-KHGKMECEPCGGLCPKACEGSGSSRYQVDDSSN 347  
 357 IGTLELFACCKKIFGSLAFPSFGDDPSNAPLQPELOVFELELTGLVLYSAMP 416  
 348 ID----GFVVCITLGNLDFLITGLVNDPNNKPLPALDEKLVFRVRELTGLNIOQSW 403  
 417 DSLPDLVSFQNLGVIRGRIIHNAYS-LTLOGIGISMLGRSLRELSGLALIHNTHLG 475  
 404 PHMNFVSFNLTLTIGRSLYNGFSLLIKHNLNLTSLGRSLKEISAGGVYISANQOC 463  
 476 FVHTVPDDQLFRPHQALHTA-NREDEDEVGSLACHQLCARGHGWGSPPTCVVCSQF 534  
 464 YHSHLWNTRLRPSSEERLDIKYDRPLGEGLAGKVCPLCSSGGGWGSPGQCLSCRNY 523  
 535 LRGOECVEBECRVLQGLPREYVNAHCLPGHPECOPNGSVTCGPGPADCVACAHYKDP 594

Db 524 SEGVGVCTHGNFLOGEPREFVHBAQCFSCHEPCLPEMEGTSTCGSGSDACACAFRQGP 583  
 QY 595 FCVARGSGVKKPDLSTMPIMKPPDEGACOPDINTHSC--VDLDDKCCPAEGQASPLT 652  
 Db 584 HCNCSCHPGLLG--AKGPIYKIPDAQNECRPHENCTGCGNGPELDCLCGQAEVMSKPH 641  
 QY 653 STVSAAVGLLVVGLVGVFGLIKRQOKIR-KYTRRLLOETELVEPLTPSGAMPNQAQ 711  
 Db 642 LVIAVTVG--LAVILMLGSGFLWYGRIGKQNRARRYLERGESIEPLDPS-ERANKVL 698  
 QY 712 MLTKETELRKVVVGSGAFGPNFVSWLVPKYSASH----LEVLRENTSPANKKEI 767  
 Db 699 ARFTKETELRKLVLSGVFG---TVHKGIWPEGESIKIPVCIKVIEDKSGRQSPQAV 754  
 QY 768 LDEAVYVAGVSGPYVRLGLICTSTVQLVTLQMLPGCLLDHVRNKGRLSGQDLNNMC 827  
 Db 755 TDMMLAVGLDHAHIVRLGLCGSSLGVLTVQYLPGLSLDHVKQKRELGLQDLNNMCV 814  
 QY 828 QIAKMSYLEDVRLVHFDLAANNVLYKSPNNHYKITDFGLARLLDDETEYHADGKVPK 887  
 Db 815 QIAKGMVYEEHSMVHRLALRVMLKSPSOVQVADLPQVADLLPDDKOLLHSEAKTPK 874  
 QY 888 WMALESILRRFTHSGDWSYGYTWELMTFGAKPYDG;PAEIPDLLEKGRLEPOPPIC 947  
 Db 875 WMALESIHGKTHSGDWSYGYTWELMTFGAEPYAGRLAIEPDLLEKGRLEPOPIC 934  
 QY 948 TTDVYIMVYKMWIDSECPREFELVSEFSRMAPDQRFVVIQNEGLDGPASPLDSTFYRS 1007  
 Db 935 TTDVYIMVYKMWIDSECPREFELVSEFSRMAPDQRFVVIQNEGLDGPASPLDSTFYRS 991  
 QY 1008 LLEDMDGLVDAEVLVYVQGFCCDPAPAGAGMWHHRSSSTRSGGDLITGLERPE 1067  
 Db 992 VLTTEL-----QEALEPEL-----DLDLLEKE 1017  
 QY 1068 E-----EAPRSLAPSE-----AGSVDFDCLMGAKAQ 1099  
 Db 1018 EGLATLSGALSALPTLTLPGRSGSLTPSSGYWMMNQSLSGEACLDVAVAGSEQFSR 1077  
 QY 1100 SLEPTDPSPLQRYSEDTVPPLSETDGYV---APL-----TC-----SPQPE---Y 1139  
 Db 1078 PLSLH-PIPRGR-----PASESSSEGHVGEAELOEKVYSYCRSRSRSPRPGDSAY 1129  
 QY 1140 VNQDPVQPQPSPREP-----LPAARPGATLERAKTISP-GKNQV----- 1181  
 Db 1130 HQQRHSLTPVPLSPGLEBEDNGVYMPDTHLKGASSRSGTSSVGLSGVTEBED 1189  
 QY 1182 KDVFARFAGAVENPEVLTPOGGAAPQFPP 1210  
 Db 1190 ED-----EEYEWNRKRKRGSP-PRPP 1209  
 RESULT 11  
 ID EGFR\_DROME STANDARD; PRT; 1426 AA.  
 AC P04412; O61601; Q9W2G0; P81868;  
 DT 13-AUG-1987 (Rel. 05, Created)  
 DT 15-DEC-1998 (Rel. 37, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Epidermal growth factor receptor precursor (BC 2.7.1.112) (Egfr)  
 DE (Gutten receptor) (Tropo protein) (Drosophila relative of ERBB)  
 GN EGFR OR TOP OR C-ERBB OR DER OR CG10079.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Preyoptera; Neoptera; Endopterygota; Diptera; Braachyera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 NCBI\_TaxId=7227;  
 RN  
 RP SEQUENCE FROM N.A. (ISOFORMS TYPES I AND II).  
 RA MEDLINE=94350209; PubMed=8070664;  
 RA Clifford R., Schubach T.;  
 RT "Molecular analysis of the Drosophila EGF receptor homolog reveals  
 RT that several genetically defined classes of alleles cluster in

RT subdomains of the receptor protein.";  
 RL Genetics 137:531-550 (1994).  
 RN [2]  
 RP REVISIONS.  
 RL Clifford R., Schubach T.;  
 RN Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 [3]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RA MEDLINE=85124611; PubMed=2982499;  
 RA Liven E., Glazer L., Segal D., Schlessinger J., Shilo B.-Z.;  
 RT "The Drosophila EGF receptor gene homolog: conservation of both  
 RT hormone binding and kinase domains.";  
 RL Cell 40:599-607 (1985).  
 RN [4]  
 RP SEQUENCE FROM N.A., CHARACTERIZATION, AND ALTERNATIVE SPLICING.  
 RC STRAIN=Oregon-R; TISSUE=Embryo;  
 RX MEDLINE=87002474; PubMed=3093080;  
 RA Schejter E.D., Segal D., Glazer L., Shilo B.-Z.;  
 RT "Alternative 5' exons and tissue-specific expression of the  
 RT Drosophila EGF receptor homolog transcripts.";  
 RL Cell 46:1091-1101 (1986).  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I), TISSUE SPECIFICITY, AND MUTATION  
 RP ANALYSIS.  
 RX MEDLINE=99102120; PubMed=9882502;  
 RA Lesokhin A.M., Yu S.-Y., Katz J., Baker N.E.;  
 RT "Several levels of EGF receptor signaling during photoreceptor  
 RT specification in wild-type, Ellipse, and null mutant Drosophila.";  
 RL Dev. Biol. 205:129-144 (1999).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM TYPE I).  
 RC STRAIN=Berkeley;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hinkins R.A., Gale R.P.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.-H., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-R.C., Blaise R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Ball R.J.F., Agbayani A., An H.-J., Andrews-Fiankoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Bencos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Borokov D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Butts K.C., Buam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferrara C., Ferreira S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,  
 RA Jaitli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., Moperson D.,  
 RA Mervinlov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheefer F., Shen H.,  
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Wolley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of Drosophila melanogaster.";  
 RL Science 287:2185-2195 (2000).  
 RN [7]  
 RP SEQUENCE OF 959-1078 FROM N.A.



```

Db      543 RPAIATQKESECKYMWVENLRADLCENKGTICSDQCNEDGCMWAGTDDCLTCNENFNNGTC 602
Qy      541 VEECNVLOGLPREYVNAHCLPCHECCQNGNSVTCFGEADQCAVAHYKDPFCVARC 600
Db      603 IADCGYISNAK--FNRTCKICHPECR-----TCNGAGDHCECHVVRDGGQCHVSEC 654
Qy      601 P-----SGVK-----PDL 608
Db      655 PKKXVDRGVRECHATCDGCTGPKDTIGACTTCNLAIIINNAIVKRLCKDKDCPD- 713
Qy      609 SYMPTWKF--PDEEGACQ-----CPI-----NCTH----- 632
Db      714 GV--FWYVHPQGGSLKPLAGRAVCRKCHPLCELCCTNYGHEQYCSKTHYKREQCET 771
Qy      633 -----SC-----VDLDKG----- 641
Db      772 ECPADHYTDEBQRECFQRHPECNGCTGPADDCCKGRNPKLPDANETGTVNSTMNCTG 831
Qy      642 -CPAECR-----ASPLTS-----IVSAVGLLVVLGVVPGI 673
Db      832 KCPLEMRHVNVOYTAIGPYCAASPPRSSKITANLDVNMIFITGAVLVPTICILCV--T 889
Qy      674 LKBPQCKIRKXT--MRRLQETELVEPTPSGAMPNQAKMLKETEELRKXKVLGSGAF 731
Db      890 YICRQKQKAKKETVKMTALSGCEDSEPLRPSNIGANLCKLIVDAELRKGGVLMGAF 949
Qy      732 GNNFTVSEFMLEVP-----KVASHLEVLRENTSPKANKELIDEAYVAVGSPYVSRL 786
Db      950 G--RVYKGVW--VPEGENVKIPVAIKELK--STGASSESEFIREAYIMASEHVALKLKL 1004
Qy      787 GICLTSTVQVLTQMPYGLLDHVENRRLSGSODLIMQMLAGMSTLEVLRLVARDL 846
Db      1005 AVCSGQMLITQMLPGLLDLYVRNRRKIGSKALLNNSTIAGMSTLEERKRLVHVDL 1064
Qy      847 AARNVLKSPNHVKITDFGLARLLDIDETEHADGKAVIKMMALESILRRFTHQSDVW 906
Db      1065 AARNVLVQTPSLVKITDFGLATLSDSNHYKAGKMKIKMLAEICRNRFTSKSDVW 1124
Qy      907 SGTAVVEMLTGAKPYDGIPLAREIPDLLEKEERLPPPICTTDVYIMVVKCMIDSECR 966
Db      1125 AGGVITWELTIGOPRHEINIPAKDIPDLLEVGLKEQPEICGLDIYCTLSGCHMDAAR 1184
Qy      967 PRFRELVSFEMARPPORFVVITQNEDELG--PASPLDSFFVYSLLEDD--DMGLVLA 1021
Db      1185 PTFKQLTTFAPAPDPGRKYLAI PGDKFTRLA-----YTSQBEKDLIRKLAFTTGS 1237
Qy      1022 EYLVPOQGFCDPAPAGAGMYHRRSSSTRSGGDLTLGLEPSEENAP-----RSL 1075
Db      1238 EAIAPKDDYLOPKAAGPS-----HRTDCT-----DEMPKLNRYCKDPS 1276
Qy      1076 ASEEGAGSVFDG---DLGMAAKGLQSLPTHDPSPLOQYSEDPTVPLPSETDGVVAPLT 1132
Db      1277 NKNSSTGDDERSSAREVGVNLR-----LDLPVDDDDYLMPT 1314
Qy      1133 CSPOPEYVQDPVRPQPPSPREGPLPAAPAGATLERAKTILSPKNGVYKDVAFAGAVE 1192
Db      1315 CQPGPNNNNNM-----NPNQNNMAAVGAAGM-----DLIGPVSVD 1353
Qy      1193 NPEYL---TPQGAAPPH-----PPAFSP-AEDNLYYMD 1224
Db      1354 NPEYLLMACTLGVGESPIPTQIGIPVMGPGTMEVYKVPMPSEPTSSDHEYYND 1408

```

```

RESULT 12
ERBB ALV
ID ERBB ALV STANDARD: PRT: 634 AA.
AC P00534;
DT 21-JUL-1986 (Rel. 01, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Tyrosine-protein kinase transforming protein erbB (EC 2.7.1.112).
V-ERBB.

```

```

OS Avian leukosis virus.
CC Viruses; Retroid viruses; Retroviridae; Alpharetrovirus.
CX NCBI_Taxid=11864;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8528222; PubMed=2988784;
RA Nilsen T.W., Maroney P.A., Goodwin R.G., Rottman P.M.,
RA Crittenden L.B., Raines M.A., Kung H.-J.;
RT "c-erb activation in ALV-induced erythroblastosis: novel RNA
RT processing and promoter insertion result in expression of an
RT amino-truncated EGF receptor.";
RL Cell 41:719-726(1985).
CC
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC
CC -I- MISCELLANEOUS: THIS PROTEIN IS PRODUCED BY ERYTHROLEUKEMIA CELLS
CC IN CHICKS WHICH HAVE BEEN INFECTED WITH THE AVIAN LEUKOSIS VIRUS
CC AND SUBSEQUENTLY DEVELOP LEUKEMIA. THE LEUKEMIAS ARE INDUCED BY
CC THE INSERTION OF RETROVIRAL PROMOTOR ELEMENTS IN OR NEAR THE
CC C-ERBB GENE, WHICH LEADS TO C-ERBB ACTIVATION.
CC
CC -I- MISCELLANEOUS: THIS PROTEIN IS SYNTHESIZED AS A GAG-ENV-ERBB
CC PROTEIN.
CC
CC -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC
CC EMBL, M10066; AAA48763.1; ALT_INIT.
CC PIR, A00643; TVCHLV.
CC PIR, B00643; TVPVLV.
CC HSSP, P11362; IFKG.
CC InterPro: IPR000719; Euk PKinase.
CC InterPro: IPR001245; Tyr PKinase.
CC Pfam, PF00069; Kinase; 1.
CC PRINTS; PR00109; TYRKINASE.
CC PRODOM; PD000001; Euk PKinase; 1.
CC SMART; SM00219; TyKc; 1.
CC PROSITE; PS00107; PROTEIN KINASE ATP; 1.
CC PROSITE; PS00109; PROTEIN KINASE TYR; 1.
CC PROSITE; PS50011; PROTEIN KINASE_DOM; 1.
CC Translase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC Glycoprotein; Phosphorylation.
CC DOMAIN 132 399
CC NP BIND 138 146 ATP (BY SIMILARITY).
CC BINDING 165 165 ATP (BY SIMILARITY).
CC ACT SITE 257 257 BY SIMILARITY.
CC SEQUENCE 634 AA; 70891 MW; E705E33A0BE01FCC CRC64;

```

```

Query Match 24.0%; Score 1638.5; DB 1; Length 634;
Best Local Similarity 50.2%; Pred. No. 8.36-81;
Matches 357; Conservative 80; Mismatches 143; Indels 131; Gaps 19;

```

```

Qy 587 CAHYKDDPPCYARPPSSVGRKDLSTYPTWKRPPDEGAQOPPINTHSCVLDKGCQAFQ 646
Db 3 CAHFDGPHCVKACPAVLGENDTL-VKTAADANAVALQHPNCTRCKGPGLEGCP- 58
Qy 647 RASPLTISVSAV--GILLVVLGVVFGILLKROQKIRKTYMRLLQETELVEPLTPSGA 705
Db 59 NGSKTPISAGVGVGLCLVVGIGLYLRR-HYRKXTRRLLOERRELVETLTPSGE 117
Qy 706 MENOQMRILKETELRKRYKVLGSGAFGPNFTVSFWLVRKVSASHLEV---LRENTSP 761
Db 118 AFNOAHRLIKETELFFKVKYVLGSGAFG---TYKKGIMPEGEKVKIPVAIKELRENTSP 173
Qy 762 KANKEILDEAYVAVGSPYVSRLGICLSTVQVLTQMLPYGLLDHVENRRLGSD 821
Db 174 KANKEILDEAYVAVSUNPHYCRLLGICLSTVQVLTQMLPYGLLDYIREHKNTSSQY 233
Qy 822 LLNMCQIANGMSTLEVDRLVHDLAARNVLYKSPNHVKITDFGLARLLDIDETEHADG 881

```



```

DB 234 LNMVCQIAKGMVYLERLVHRDLAARVILVKTQHVKITPTFGIAKLGNBEXHYHEG 293
QY 882 GXYPIKMALESILRRRFTHQSDVSYGVTVWELMTFGAKKPYDGIIPAREIPDLKEGERL 941
DB 294 GXYPIKMALESILRRRFTHQSDVSYGVTVWELMTFGSKPYDGIIPAREIPDLKEGERL 353
QY 942 POPPCTIDVMMVYCMVIMDESECRPELSESRMARDQRFVVIQ-NEDLGPASPL 1000
DB 354 POPPCTIDVMMVYCMVIMDESECRPELSESRMARDQRFVVIQ-NEDLGPASPL 413
QY 1001 DSTFYSLLDDMDGLVDAEYLVPOQGFPCPDPAAGAGVHHRHSSSTRSGGDLT 1060
DB 414 DSKEFYRTLMEEEMEDIVDAEYLVHOGFF-----NSPST----- 449
QY 1061 LGLEPSEEARSPUL-----APSEGSQVDFPDGLGMAKGLQSLPTDPSPLQRYSED 1115
DB 450 -----SRPLSLSLATSNNSATNCID-----RNGQGHVREDSEFVQRYSSD 491
QY 1116 PTVPLPSET--DGYVAPLTCSPQPEYVNOQDVRPOPSPREBPLPAARPAATLERATL 1173
DB 492 PTGNFLESIDGFL-----PAPEYVNO--LMPKXPS-----T 522
QY 1174 SPQKNGVAVDF-----AFGAVNPEVITTPQGAAPQHPAPSPAPD 1218
DB 523 AVVQNIYNNISLTAISKLPMSRYONSHSTAVNPEYL-----NTNQSPLAKTYFE 574
QY 1219 NUYVMDQ-----DPE-----RGAPSTFKGTPTAENPEYGLDVP 1254
DB 575 SSFYWTQSGNHQINLDNPDYQDFLENERKPNGLKVPAAENPEYLRYVAP 625

RESULT 13
EGFR CHICK
ID EGFR CHICK STANDARD; PRT; 703 AA.
AC P1387;
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Epidermal growth factor receptor precursor (EC 2.7.1.112) (CER)
DE (Fragment).
GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
ON NCBI_TaxID=9031;
RX SEQUENCE FROM N.A.
RP MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Howk R., Sap J., Bellot F., Winkler M.,
RA Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT transforming growth factor alpha, and differential binding of EGF and
RT expression in mouse cells, and differential binding of EGF and
RT Mol. Biol. 8:1970-1978 (1988).
RL FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - SUBCELLULAR LOCATION: Type I membrane protein.
CC - MISCELLANEOUS: Binding of EGF to the receptor leads to
CC dimerization, internalization of the EGF-receptor complex,
CC induction of the tyrosine kinase activity, stimulation of cell DNA
CC synthesis, and cell proliferation.
CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.jeb-sib.ch/announce/
CC or send an email to license@jeb-sib.ch).
CC -----
DB EMBL; M20386; AAA48760.1; -
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; EGF_kinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00757; Furin-like; 2.
DR SMART; SM00261; FU; 4.
DR PROSITE; PS00107; PROTEIN KINASE_TYR; PARTIAL.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; PARTIAL.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; PARTIAL.
KW Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; phosphorylation.
FT SIGNAL 1 30
FT CHAIN 1 30
FT DOMAIN 31 654
FT TRANSMEM 655 667
FT DOMAIN 668 >703
FT DISULFID 197 206
FT DISULFID 201 214
FT DISULFID 222 230
FT DISULFID 226 238
FT DISULFID 239 247
FT DISULFID 243 255
FT DISULFID 258 267
FT DISULFID 271 298
FT DISULFID 302 314
FT DISULFID 318 333
FT DISULFID 336 340
FT DISULFID 513 522
FT DISULFID 517 530
FT DISULFID 533 542
FT DISULFID 546 562
FT DISULFID 565 581
FT DISULFID 569 589
FT DISULFID 592 601
FT DISULFID 605 627
FT DISULFID 630 638
FT DISULFID 634 646
FT CARBOHYD 134 134
FT CARBOHYD 190 190
FT CARBOHYD 200 200
FT CARBOHYD 359 359
FT CARBOHYD 368 368
FT CARBOHYD 420 420
FT CARBOHYD 573 573
FT CARBOHYD 578 578
FT CARBOHYD 613 613
FT CARBOHYD 633 633
FT CARBOHYD 648 648
FT NON_TER 703 703
SQ SEQUENCE 703 AA; 77427 MW; AFF2DE1B735A690 CRC64;

Query Match 23.4%; Score 1595; DB 1; Length 703;
Best Local Similarity 44.6%; Pred. No. 2e-78;
Matches 316; Conservative 111; Mismatches 251; Indels 30; Gaps 14;

QY 8 RMGLALLLPFGAA-----STGYCTGDMKRLPASPEHLDMLRLHYGCGVYQGNLE 61
DB 13 RGAIVLVLLVLLGVALGSAVEBKVCQGTNNKRLQLGHEVDFHPSLQRMVNNCEVVSNTLE 72
QY 62 LTYPTASASFLDIDIEVGYVLIANQVRVPLQRLRVRTGQLFEDNYALAVLNGD 121
DB 73 ITVEYHNRDLTFLFKTIDEVAGVYLIANMVDVPLELQIIRGVLYVDSNFALAVSNYH 132
QY 122 PLNNTVTGASPCGLRELDLRSTLTILKGVYIQRPOCYOPTIMKPIFKNQDLAL 181
DB 133 -MNTQ-----GLRELPMKRLSELINGVATSNPKLCNMDYVAMNDIITSKR-PL 182
QY 182 TLID-TNRSPACHPCSPMCKGSRGWSESSDCSLRTVYAGGCA-RCKGPLEPTDCHEQ 239

```

```

Db      183 TVLDFASNLSSCPKCHNCTEDHOCWAGBONCOTLTKVICAQCCSRCKGKVPSCCHNQ
Qy      240 CAAGCTGPHSDCLACLFHNSGICELHPALVYNTDTPESMPNDEGGTYGASCVTAC
Db      243 CAACTGCPRESDDLAKRFRDQATCQDTPRLVYNTPTTYQMDVNPBGKTSGATCVREC
Qy      300 PNYVLTSDVSCCTLVCPHNOEYTAEDGTORCEKSKPCARVYCYGLMGVYIKANSFTGI
Db      303 PHNYVLTSDVSCCTLVCPHNOEYTAEDGTORCEKSKPCARVYCYGLMGVYIKANSFTGI
Qy      360 TELE-PAGCKKIFGSLAFLPESPDGPAANTAPLOEQOVETLEIETGLYLIAMPS 418
Db      361 TNDISFRKCTKINGDVSLIPVAFGLAFLKTPLODKLDVFRVYEISGFLLIQAMPBN 420
Qy      419 LPDLVFNQLQVIRGRILHNGAVSLTLQGLISWLSRLSRELSGSLALIHNTLCPFH 478
Db      421 ATDLVAFENLEIRGTRKQHQGLSLAVNMLKIQSLGRLKEISDDIIMKNKULCVAD 480
Qy      479 TVWMDOLFRRPHOALHTANRPDECEVGGGLACGLCAAGHCWGPPTQCVNCSQFLRQ 538
Db      481 TMMRSLEFATQSGTKIKIONRNKNDCTADRHVCDPLCSVGCWGPFPFHCSCRFESRQK 540
Qy      539 ECVEECRVQLQGLPREYVNAHCLPCHEPCOPQNG--SVTCFGEPAQCCVCAHYKDPF 595
Db      541 ECVKQCNIIQGEFRERERDSKCLPCHSECLVQNSTAYNTGSGPGRDHCKMAHFDGPH 600
Qy      596 CVARCPGVKPDLSYPIWKFPEDEGACQPCPINTCHSCVDDDKCPAEGASPLTSIV 655
Db      601 CVACCPAGVLGENDTL-VMKYADANAVCOLCHPNCTRGCKGFLGEGCP---NGSKTPSIA 656
Qy      656 SAVV-GILVVVLGVVFGILIKRQOKIRKYMRLLOTELEVEPLTP 702
Db      657 AGVVGSLCLVVGGLGLGLYLRNR-HIVKRTILRLQRELEVEPLTP 703

```

```

CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      EMBL, K02006; AAA42394.1; ALT. INIT.
CC      EMBL, K01216; AAA42400.1; -.
CC      PIR, A00644; TYVTH.
CC      HSSP, P11362; 1BCK.
CC      InterPro, IPR000719; Euk_pkinase.
CC      InterPro, IPR001245; Tyr_pkinase.
CC      Pfam, PF00069; Pkinase.1.
CC      Prodom, PD000001; Euk_pkinase; 1.
CC      SMART, SMO0219; TyrKc; 1.
CC      PROSITE, PS00107; PROTEIN_KINASE_ATP; 1.
CC      PROSITE, PS00109; PROTEIN_KINASE_TYR; 1.
CC      PROSITE, PS00011; PROTEIN_KINASE_DOM; 1.
CC      Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;
CC      Glycoprotein; Phosphorylation.
CC      DOMAIN 132 399
CC      NP_BIND 138 146 ATP (BY SIMILARITY).
CC      BINDING 165 165 ATP (BY SIMILARITY).
CC      ACT_SITE 257 257 BY SIMILARITY.
CC      CONFLICT 29 29 R -> W (IN REF. 2).
CC      CONFLICT 140 140 S -> F (IN REF. 2).
CC      CONFLICT 146 146 I -> V (IN REF. 2).
CC      SEQUENCE 604 AA; 67633 MW; 76BCDD06745D609 CRC64;

Query Match 23.4%; Score 1592; DB 1; Length 604;
Best Local Similarity 50.0%; Pred. No. 2.4e-78;
Matches 347; Conservative 77; Mismatch 136; Indels 134; Gaps 18;

Qy      587 CAHYKDPFPCVAPCPGPDLSYPIWKFPEDEGACQPCPINTCHSCVDDDKCPAEG 646
Db      3 CAHFDGPHCVKACPAVLGENDTL-VKRYADANAVCOLCHPNCTRGCKGFLGEGCP--- 58
Qy      647 RASPLTSIVSAV-GILVVVLGVVFGILIKRQOKIRKYMRLLOTELEVEPLTPSGA 705
Db      59 NGSKTPSIAAVVVGGLCLVVGGLGLYLRNR-HIVKRTILRLQRELEVEPLTPSGE 117
Qy      706 MPNOAQMILKETELRKVKVLGSGAFGNNFTVSFWLVRPKVASHLEV---IRENSP 761
Db      118 APNOQHMLKILETEPKKVLGSGAFG---TIKGLMTEPEGEVKVLPVAIKELREATSP 173
Qy      762 KANKEIIDEAYVAVGSPVYRLIGICTSTVQVLTQLMFYGGLDHPVRENRGLSGOD 821
Db      174 KANKEIIDEAYVAVSVDPHVCRLGLICTSTVQVLTQLMFYGGLDHPVRENRGLSGOY 233
Qy      822 LIAHNCQIAKMSLIEDVRLVHRLAARVLYKSPNHYKTKDIFGLARLDDIEFYHAGD 881
Db      234 LIAHNCQIAKMSLIEDVRLVHRLAARVLYKSPNHYKTKDIFGLARLDDIEFYHAGD 881
Qy      942 POPPCTIDVYIMVKKCMIDSECRPREIVSEFSRMAQDPQFVVIQ-NEDIGPASPL 1000
Db      354 POPPCTIDVYIMVKKCMIDSECRPREIVSEFSRMAQDPQFVVIQ-NEDIGPASPL 1000
Qy      1001 DSTFYRLIEDDDMDGLVDAEYVLPVQGFCCPRPAGAGAMVHRRSSSTRGGGDLT 1060
Db      414 DSKFYRLIEDDDMDGLVDAEYVLPVQGFCCPRPAGAGAMVHRRSSSTRGGGDLT 1060
Qy      1061 LGLPSEBEARSP-----APSEGAQDFDGLGMAKGLSLPTPHSPLOQRYSED 1115
Db      450 -----SRTPLSLSLSTSNNSATNCID-----RNGGHVREDSEFYQRYSSD 491
Qy      1116 PTVLPLPSET-DGVAVLTCGPOPEYVNOQVPRQPSRREGPLPAARPAQATLERAKTL 1173

```

DB 492 PTGNFLEESIDDEFL-----PAPEVYNQ--LMPKXSTAM----- 524

QY 1174 SPKNGVYKDVFAF-----GGAVERNPEYTPQGAAPQPHPPAPS 1214

DB 525 -----VQNOIYNFISLTAISKLPMSRYONSHTADNPEYL-----NTQSPFLAK 570

QY 1215 PAEDNLVYWDQDPERGAPSTFKGTPTAENPEY 1248

DB 571 TVFESSPYWIOGNNHQ-----INLDNPDY 594

RESULT 15

ERBB\_AVIEW STANDARD; PRT; 540 AA.

ID ERBB\_AVIEW

AC P1273;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Tyrosine-protein kinase transforming protein erbb (EC 2.7.1.112).

GN V-ERBB.

OS Avian erythroblastosis virus (strain ts167).

OC Viruses; Retroid viruses; Retroviridae; Avian type C retroviruses.

OX NCBI\_TaxID=103898;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=67064456; Pubmed=2878364;

RA Choi O.R., Trainor C., Graf T., Beug H., Engel J.D.;

RT "A single amino acid substitution in v-erbB confers a thermolabile phenotype to ts167 avian erythroblastosis virus-transformed erythroid cells."

RT Mol. Cell. Biol. 6:1751-1759(1986).

CC -1- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein tyrosine phosphate.

CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M13179; AAA42401.1; -

DR PIR: A25231; TVFVEB.

DR HSSP: P11362; 1EGK.

DR InterPro: IPR000719; Euk\_pkinase.

DR InterPro: IPR001245; Tyr\_pkinase.

DR Pfam: PF00069; pkinase; 1.

DR ProDom: PD000001; Euk\_pkinase; 1.

DR SMART: SM00219; TyrKc; 1.

DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.

DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.

DR PROSITE: PS50011; PROTEIN KINASE DOM; 1.

DR Transferase; Tyrosine-protein kinase; ATP-binding; Oncogene;

KW Glycoprotein; Phosphorylation.

FT DOMAIN 132 399

FT NP BIND 138 146

FT BINDING 165 165

FT ACT SITE 257 257

FT VARIANT 270 270

FT SEQUENCE 540 AA; 60412 MW; 5B53297A068B65D CRC64;

Query Match 22.4%; Score 1524; DB 1; Length 540;

Best Local Similarity 52.8%; Pred. No. 9.3e-75;

Matches 329; Conservative 70; Mismatches 126; Indels 98; Gaps 16;

QY 587 CARYKDPFVACPSGVKDLVMPWKFPDEGACQPPINCTHSYVDLDDKGPAPQ 646

DB 3 CAHFIDGPHCVKACPGAVGENDTL-VWKYADANAVCQLCHPNCRTGCKGPGLEGCP--- 58

QY 647 RASPLTISVAVV-GILLVVVLGVFGILIKRQOKIRKTYMRLLQETELVPLTPSGA 705

DB 59 NSKTPSIAGVVGGLCLVVGIGLYLRR-HYVRKRTLRLQERLVEPLTPSGE 117

QY 706 MPNOQAMBILKETELRKXKVLGSAFGNNFTVSFWLRVPKVSASHLEV-----LRENTSP 761

DB 118 APNOAHILKLETERKXKVLGSAFG---TVKGLMIPEGEKVTTFVAIKELREATSP 173

QY 762 KANKEILDEAVYVAVGSPPVYRLLGLCLTSVQLVTVQJLMEYGLLDHVRNRRGLSGD 821

DB 174 KANKEILDEAVYVAVSVDPNPHVCRLLGLCLTSVQLVTVQJLMEYGLLDHVRNRRGLSGD 233

QY 822 LLNMCQIAKGNSTYEDVRLVHRDLAANVLYKSPNRYKTFDEGLARLLDDETEYHAGD 881

DB 234 LLNMCQIAKGNSTYEDVRLVHRDLAANVLYKTPQHVKTIDFGLAQGLADEKEYNAGG 293

QY 882 GKVPKXMALESILRRRTTHSDVWYSGVTWYELMTFGAKPYDGIAPAREIDLEKGERL 941

DB 294 GKVPKXMALESILHRTYTHSDVWYSGVTWYELMTFGAKPYDGIAPAREIDLEKGERL 353

QY 942 POPPICTIDVYMIWYKCMWIDSECPREFRELVSFESRMARDPQPFVVIQ-NEDLGPAAPL 1000

DB 354 POPPICTIDVYMIWYKCMWIDSECPREFRELVSFESRMARDPQPFVVIQ-NEDLGPAAPL 413

QY 1001 DSTFYSRLDEDDMDGLVDABEYLVPQGFCCPPAPAGAGVHRRSSSTRSGGGLT 1060

DB 414 DSKFYRTLMSEEDMEDIVDADEYLVPHQGF-----NSPST----- 449

QY 1061 LGLPSEBEAPRSP-----APSEGAGSDVECDGLGMAKAGLQSLPTHPDSPLOQRYSED 1115

DB 450 -----SRTPLSLSLATSNNSATINCIDRNG-----H----- 476

QY 1116 PTVPLPSTDGVVAFLTSPQPEYVNOQDVAPQPPSPREGLPAARAPGAT-LEBAKTL 1174

DB 477 -----PVREDGFL-----PAPEVYNQ--LMPKXSTAMVQNIYVISTLAIKLPIDS 523

QY 1175 PCKNGVYKDVFAFGAVENPEYL 1197

DB 524 RYQN-----SHSTAVDNPEYL 539

Search completed: July 22, 2003, 09:19:00

Job time : 21.642 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 22, 2003, 09:01:42 ; Search time 46.8246 Seconds

(without alignments)  
5522.503 Million cells updates/sec

Title: SEQ4-710-730-14

Perfect score: 6813

Sequence: 1 METALCRWGULLALLPPGA.....TFKGTPTANEPEYGLDVPV 1255

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: SP\_ARCHA:.\*  
2: SP\_BACTERIA:.\*  
3: SP\_FUNGI:.\*  
4: SP\_HUMAN:.\*  
5: SP\_INVERTEBRATE:.\*  
6: SP\_MAMMAL:.\*  
7: SP\_MHC:.\*  
8: SP\_ORGANELLE:.\*  
9: SP\_PHAGE:.\*  
10: SP\_PLANT:.\*  
11: SP\_RODENT:.\*  
12: SP\_VIRUS:.\*  
13: SP\_VERTEBRATE:.\*  
14: SP\_UNCLASSIFIED:.\*  
15: SP\_VIRUS:.\*  
16: SP\_BACTERIAP:.\*  
17: SP\_ARCHA:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	6091	89.4	1259	6	018735
2	3031	44.5	1209	1	03QX70
3	3004	44.1	1210	11	03EP98
4	2617	38.4	1165	13	03YH40
5	2580.5	37.9	1137	13	03W666
6	2204	32.3	1328	13	P79754
7	1923.5	28.2	1433	5	09B1H9
8	1871	27.5	419	4	09UK79
9	1739	25.5	367	11	08R2X1
10	1697.5	24.9	412	15	08WYV0
11	1609	23.6	729	15	086712
12	1607	23.6	567	15	086714
13	1547.5	22.7	562	15	084895
14	1539	22.6	545	15	085468
15	1506.5	22.1	655	11	09WV65
16	1490.5	21.9	643	11	09ERV6

17	1201	17.6	1193	5	09Y1X8	09Y1X8 ephydria f
18	1126	16.5	527	13	090836	090836 gallus gall
19	1117.5	16.4	1368	5	023821	023821 caenorhabdi
20	1114	16.4	1717	5	026566	026566 schistosoma
21	1001.5	14.7	478	11	09ESE0	09ESE0 rattus norv
22	942.5	13.8	599	13	09PSH2	09PSH2 gallus gall
23	906	13.3	165	4	014256	014256 homo sapien
24	887	13.0	176	11	0923V5	0923V5 rattus norv
25	806.5	11.8	346	13	P11776	091176 xiphophorus
26	778	11.4	435	5	08S2M1	08S2M1 xiphophorus
27	754.5	11.1	311	13	099162	099162 xiphophorus
28	734	10.8	331	4	09BUD7	09BUD7 homo sapien
29	723	10.6	149	6	09BC66	09BC66 oryzocolagus
30	695.5	10.2	1362	13	09PVZ4	09PVZ4 xenopus lae
31	679	10.0	1671	5	09NVJ5	09NVJ5 biophthalari
32	655	9.6	1418	13	093457	093457 scophthalmu
33	646.5	9.5	1368	13	08UW85	08UW85 paralicthy
34	631.5	9.3	1369	13	08UW86	08UW86 paralicthy
35	630.5	9.3	1472	5	09U5A8	09U5A8 bombyx mori
36	621	9.1	1412	13	08UW84	08UW84 paralicthy
37	611	8.9	1358	13	073798	073798 xenopus lae
38	604.5	8.9	1418	13	08UW83	08UW83 paralicthy
39	592	8.7	1245	13	09YGH8	09YGH8 scophthalmu
40	583	8.6	1371	11	09QVW4	09QVW4 rattus sp.
41	580.5	8.5	2144	5	09VD94	09VD94 dirosophila
42	578.5	8.5	1091	4	09UW04	09UW04 homo sapien
43	572	8.4	987	11	091YMO	091YMO mus musculu
44	569	8.4	935	4	096L35	096L35 homo sapien
45	569	8.4	987	11	099WR2	099WR2 mus musculu

## ALIGNMENTS

## RESULT 1

ID	018735	PRELIMINARY;	PRT;	1259 AA.
AC	018735;			
DT	01-JAN-1998 (TREMBLrel. 05, Created)			
DT	01-JAN-1998 (TREMBLrel. 05, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Erpb-2.			
OS	Canis familiaris (Dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
OX	NCBI_TaxId=9615;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Yokota H.;			
RT	"cDNA cloning of erpb-2 from canine mammary gland."			
RL	Submitted (OCT-1997) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AB008451; BAA23127.1; ..			
DR	HSSP; P11362; IFGK.			
DR	InterPro; IPR002048; EF-hand.			
DR	InterPro; IPR000494; EGFR_L domain.			
DR	InterPro; IPR000719; Euk PKinase.			
DR	InterPro; IPR002174; Furin-like.			
DR	InterPro; IPR001245; Tyr PKinase.			
DR	InterPro; IPR004019; YIP_motif.			
DR	Pfam; PF00757; Furin-like; 1.			
DR	Pfam; PF00069; pkinase; 1.			
DR	Pfam; PF01030; Recep_L domain; 2.			
DR	Pfam; PF02757; YIP_2.			
DR	ProDom; PD000001; Euk_pkinase; 1.			
DR	SMART; SM00261; Fy3.			
DR	SMART; SM00219; Tyrc; 1.			
DR	PROSITE; PS00018; EF_HAND; UNKNOWN; 1.			
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.			
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.			
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.			
KW	ATP-binding; Transferase; Tyrosine-protein kinase.			
SQ	SEQUENCE 1259 AA; 137969 MW; E37364D4C4ACD66 CRC64;			

Query Match 89.4%; Score 6091; DB 6; Length 1259;  
 Best Local Similarity 89.2%; Pred. No. 0;  
 Matches 1188; Conservative 49; Mismatches 72; Indels 16; Gaps 5;

QY 1 METALACRGLLALLPPGAASQVCTGDMKLRLEPASETHLDMRLHYQCCQVYQGL 60  
 DB 1 METALACRGLLALLPPGAASQVCTGDMKLRLEPASETHLDMRLHYQCCQVYQGL 60  
 QY 61 ELTYLPANASLSLQIQEVQGVLIHNSQVQIPIQRLIRYRGQLFEDNYALVLDNG 120  
 DB 61 ELTYLPANASLSLQIQEVQGVLIHNSQVQIPIQRLIRYRGQLFEDNYALVLDNG 120  
 QY 121 DPLNNTTPTVAGSPGLRELRSLTEILKGVLIQRNPOLCYDTIWKDIFKKNQIA 180  
 DB 121 DPLNNTTPTVAGSPGLRELRSLTEILKGVLIQRNPOLCYDTIWKDIFKKNQIA 180  
 QY 121 DPLNNTTPTVAGSPGLRELRSLTEILKGVLIQRNPOLCYDTIWKDIFKKNQIA 180  
 DB 121 DPLNNTTPTVAGSPGLRELRSLTEILKGVLIQRNPOLCYDTIWKDIFKKNQIA 180  
 QY 181 LTLIDTNRBACHPCSPMCKGSRGWSESDCCSLTRTCAGGACRCKEPLTDCHEQC 240  
 DB 181 LTLIDTNRBACHPCSPMCKGSRGWSESDCCSLTRTCAGGACRCKEPLTDCHEQC 240  
 QY 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTPESMPNDEGRYTFGASCVTACP 300  
 DB 241 AAGCTGPKASDCLACHFNHSGICELHCPALVTYNTDTPESMPNDEGRYTFGASCVTACP 300  
 QY 301 YNYLSTNDSCTLYVCELIHQEYTAEDGTQCEKSCPCARVCYGLMOYIKANSKFIGIT 360  
 DB 301 YNYLSTNDSCTLYVCELIHQEYTAEDGTQCEKSCPCARVCYGLMOYIKANSKFIGIT 360  
 QY 361 ELTEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVFETLEITGLYISAMPDILP 420  
 DB 361 ELTEFAGCKKIFGSLAFLPESFDGDPASNTAPLOPEOLQVFETLEITGLYISAMPDILP 420  
 QY 421 DLSVFONLOVIRGRIIHNGAYSITLQGLISWGLRSLRELCSGIALIHNTLHCPVHTV 480  
 DB 421 DLSVFONLOVIRGRIIHNGAYSITLQGLISWGLRSLRELCSGIALIHNTLHCPVHTV 480  
 QY 481 PMDOLFRNPHOALLHTANRPEDECEGELACHQLCARHGMCPGPTQCNCSOPLRGRC 540  
 DB 481 PMDOLFRNPHOALLHTANRPEDECEGELACHQLCARHGMCPGPTQCNCSOPLRGRC 540  
 QY 541 VEECRVLOGLPREYVNAHCLPCHPECPQONGSVTCFGBEADQVACAHYKDPFCVARC 600  
 DB 541 VEECRVLOGLPREYVNAHCLPCHPECPQONGSVTCFGBEADQVACAHYKDPFCVARC 600  
 QY 601 PSQVPRDLSYMPYKWPDEEGACOPCINCTHSCVDLDKGGPACQASPLTSYSAVVG 660  
 DB 601 PSQVPRDLSYMPYKWPDEEGACOPCINCTHSCVDLDKGGPACQASPLTSYSAVVG 660  
 QY 660 ILLVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQWRIKETEL 720  
 DB 660 ILLVVLGVVFGILIKRQOKIRKTYMRLLQETELVEPLTPSGAMPNQAQWRIKETEL 720  
 QY 721 RKVKULGSAFGFNNFTVFWLRP----KVSASHLEVLRENTSPKANKELLDAAYMA 775  
 DB 721 RKVKULGSAFGFNNFTVFWLRP----KVSASHLEVLRENTSPKANKELLDAAYMA 775  
 QY 775 GVSAPVSRLLGLCTSTVOLTPMPYGLLDHYRENGRSGSODLNMCMQIAKMSY 835  
 DB 775 GVSAPVSRLLGLCTSTVOLTPMPYGLLDHYRENGRSGSODLNMCMQIAKMSY 835  
 QY 835 LEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPYKMALESIL 895  
 DB 835 LEDVRLVHRDLAARNLVKSPNHVKITDFGLARLLDIDETEHADGKVPYKMALESIL 895  
 QY 895 RRRFTHQSDVMSGVTVVWELMTFGAKPYDGIPIARELPDLLEGEELPOPCITIDVYIM 955  
 DB 895 RRRFTHQSDVMSGVTVVWELMTFGAKPYDGIPIARELPDLLEGEELPOPCITIDVYIM 955  
 QY 955 VECWMDISECRPRFRELVEFSRMAADPQRFVYVQNEGLSPASPLDSTFYRSLLEDDMG 1015  
 DB 955 VECWMDISECRPRFRELVEFSRMAADPQRFVYVQNEGLSPASPLDSTFYRSLLEDDMG 1015  
 QY 1015 DLVDAEVLVPOQGFPCDPDAPGAGCMYHHRSSSTRGGGDLTLGLEPSEBEAPRSPL 1075  
 DB 1015 DLVDAEVLVPOQGFPCDPDAPGAGCMYHHRSSSTRGGGDLTLGLEPSEBEAPRSPL 1075

DB 1015 DLVDAEVLVPOQGFPCDPDAPGAGCMYHHRSSSTRGGGDLTLGLEPSEBEAPRSPL 1075  
 QY 1075 DLVDAEVLVPOQGFPCDPDAPGAGCMYHHRSSSTRGGGDLTLGLEPSEBEAPRSPL 1075  
 QY 1075 DLVDAEVLVPOQGFPCDPDAPGAGCMYHHRSSSTRGGGDLTLGLEPSEBEAPRSPL 1075  
 DB 1075 DLVDAEVLVPOQGFPCDPDAPGAGCMYHHRSSSTRGGGDLTLGLEPSEBEAPRSPL 1075  
 QY 1135 APSEGASDVFDGLGNGAAKGLQSLSTHPSPLQKRSBPYPLPSETGCVAPLTCSP 1135  
 DB 1135 APSEGASDVFDGLGNGAAKGLQSLSTHPSPLQKRSBPYPLPSETGCVAPLTCSP 1135  
 QY 1136 QPEYVNPDPVAPQPPSPREGPLPAAPAGATLER-----AKTSPGKGVKQVFAFGA 1190  
 DB 1135 QPEYVNPDPVAPQPPSPREGPLPAAPAGATLER-----AKTSPGKGVKQVFAFGA 1190  
 QY 1191 VENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPERGAPSTKGTPTAENPEYL 1250  
 DB 1195 VENPEYLTPOGGAAPQHPPPAFSPAFDNLVYWDQDPERGAPSTKGTPTAENPEYL 1254  
 QY 1251 LDVPU 1255  
 DB 1255 LDVPU 1259

RESULT 2  
 Q9OX70 PRELIMINARY; PRT: 1209 AA.  
 ID Q9OX70  
 AC 01-MAY-2000 (Tremblrel. 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor.  
 GN EGFR.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RX MEDLINE=9025868; PubMed=2342466.  
 RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,  
 RA Earp H.S.;  
 RT "A truncated, secreted form of the epidermal growth factor receptor is  
 RT encoded by an alternatively spliced transcript in normal rat tissue";  
 RL Mol. Cell. Biol. 10:2973-2982(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RA Petch L.A.;  
 RL Submitted (NCV-1991) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FISHER; TISSUE=LIVER;  
 RA Guttridge K., Dawson T.L., Earp H.S.;  
 RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.  
 DR EMBL: M37394; AAF14008.1; -.  
 DR HSSP: P13362; 1FGK.  
 DR InterPro: IPR000494; EGFR\_L domain.  
 DR InterPro: IPR000719; Euk\_Dkinase.  
 DR InterPro: IPR001245; Tyrc\_Pkinase.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF00069; Pkinase; 1.  
 DR Pfam: PF01030; Recep\_L domain; 2.  
 DR PRINTS: PRO0109; TYRKINASE.  
 DR ProDom: PD000001; Euk\_Pkinase; 1.  
 DR SMART: SMO0261; Fu\_3.  
 DR SMART: SMO0219; Tyrc; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00101; PROTEIN\_KINASE\_DOM; 1.  
 DR APC-binding; Receptor; Tyrosine-protein kinase.  
 KW SEQUENCE 1209 AA; 134891 MW; 966EERF6C1B773 CRC64;

Query Match 44.5%; Score 3033; DB 11; Length 1209;  
 Best Local Similarity 49.1%; Pred. No. 3; 7e-223;

Matches, 630; Conservative 166; Mismatches 362; Indels 124; Gaps 30;

QY 3 LALRCWGLLALLPFGA-ASTQVCTGTMKRLPASPETHLMDLRLHLYOGQVQVGNLE 61  
 Db 15 LALCAAG-----GALBKKVCGGTNSRLTQCTFEDHLSLQRMNNEEVLGNLE 66  
 QY 62 LTYLPNNASISFQDIQEVQGYVLIHNOVROVPLORLIVKGTQLEFNNYALAVDNGD 121  
 Db 67 ITVQNRNDLSFKTIQEVAGYVLIHNTVERIPLENIQIIRGNALYENTYALAVLSN-- 124  
 QY 122 PLNNTPVTGASGPGRELORLSLTEILKGVLIQNRNPQCODTLTMDIFHKNQAL 181  
 Db 125 -----YGTNTGTGRLPMRLQIILGNARFNNPILCMETIQMDIV-QDVFLSN 175  
 QY 182 TLIDITNRS-RACHPCSPCKGSRFCWSESSEDCQSLTRVACAGCA-RCKGPLPTDCHEQ 239  
 Db 176 MSMDVQRHLTGCKPCDPCNGSCMGCEENCKLTKIICAQCCSRCKGRSPDCCNQ 235  
 QY 240 CAAGCTGPGKSDCLAHFNHSGICELHCBALVTYNTDFESKNPEGRYTGASGVTAC 299  
 Db 236 CAAGCTGPGKSDCLAHFNHSGICELHCBALVTYNTDFESKNPEGRYTGASGVTAC 295  
 QY 300 PNYVLTSDVSCCTVCPLEHNOEVTADGTQRCCKSPCARVCGYIGMOYIANSKFTIGI 359  
 Db 296 PRNYVVTDHGSCVRACGPDYEV-EEDGVSKCKKCGPCRYKONGIGICEFK-DTISIVA 353  
 QY 360 TELE-FAGCKKITGSLAFIPESFDGPAANTAPLOEQQVETLEITGYLYISAMPDS 418  
 Db 354 TNKHFKYCAISGDIHILPVAFKGSFRTPELDRLEIITKTYEIRIGFLLIQAWEEN 413  
 QY 419 LPDLSFQNLQVIRGRILHNGAVSLTLOGISWLSRLSRLSGLALIHNTHLCFPH 478  
 Db 414 WDLHAFENLEIRGTQKQHGQSLAVGELNITSLGRSLKESDSDVVISGNRLCYAN 473  
 QY 479 TVPMDOLFERNPHOALHTANRPEDECVGGLACHQICARHGCWGPPTQVNCOSQFLRQ 538  
 Db 474 TTNMKLFGTPNOKTKIMNRAKDKATNHCNPLCSSEGCPEPTDVCQVWRSRGR 533  
 QY 539 ECEEGCRVLOGLPREVYNARHCLPCHPECOPOGSGTTCGPFENDQVAAAHKDDPFCVA 598  
 Db 534 ECVDKNLIEGEERREVENSECTQCHPECLPOTMTTTCGRGPDNITKCAHVVDGHCYK 593  
 QY 599 RCPSSGVKPLSYPIWKPFDEBACQPCPINCTHSCVDLDDKCPAECQASP-LTISVA 657  
 Db 594 TCSSGIMGENNTL-VMKFADANNVCHLCHANCTYGCAGGLKGC-QQPEPKISISLNG 650  
 QY 658 VGCILLVWVGVVFGI-LIKRQOKIRKTYMRLLDETLEVLPTPGSGMENVQAQRIK 716  
 Db 651 IVGGILFTIV-VALGIGLFMRRLQVRRKTLRLLELRELVLEPLTPSGEAPQAHRLIK 709  
 QY 717 ETELKRVKYLGSAGFNNFTVSFMLRVFKVASHLEV-----LRENTSPKANKELIDEAY 772  
 Db 710 ETEFKIKYLGSAFG-----TYKGLMIPGEKVKI PVALIKELREITSKANKELIDEAY 765  
 QY 773 VVAGVSPYVSRLLIGILTSTVQVLTQMPYCGCLDVEENRGLSGQDLNWCQIANG 832  
 Db 766 VVASVDPNPHVCRLGICLTSTVQVLTQMPYCGCLDVEENRGLSGQDLNWCQIANG 825  
 QY 833 MSYLEDVRLVHRDLARVNLVSPNPKITDFGLARLLDIDETEVHADGKXPIWMAL 892  
 Db 826 KMYLEDRLVHRDLARVNLVKTPOHVKITDGLALDLAEEETHAEGKXPIWMAL 885  
 QY 893 SLRRRFTQSDVWSYGVTVWELMTFGAKPYDGIAPAREIPDLLEKGERLPQPICTIDY 952  
 Db 886 SLHRIYTHQSDVWSYGVTVWELMTFGSKPYDGIAPAREIPDLLEKGERLPQPICTIDY 945  
 QY 953 MIMVKKMIMDSECRPFRELVSFMSMADPQRFVITQ-NEDLGRASPDSFTFYSLED 1011  
 Db 946 MIMVKKMIMDSECRPFRELVSFMSMADPQRFVITQ-NEDLGRASPDSFTFYSLED 1005  
 QY 1012 DMGDLVDAEEYLVPOQGFCCPDPAFAGAGMHHRRSSSTSGGDLTLGLEPSEEAR 1071  
 Db 1006 EDMEDVDADEYLVPOQGF-----NSPST-----S 1031

QY 1072 RSLPASEGAGSDYFDGDLGAAKGLQSLPTHDPSPLOXYSEDPTVPLPSET--DGVA 1129  
 Db 1032 RPLSLSSISANSN-----SSTVACINRNGSCRVKEDAFLOXYSDPTSVLTEDNIDDFL- 1086  
 QY 1130 PLTCSPOPEYNYQDVARPQPSRREGPLPAPARPAGATLERAKTISPCKGVYKVFYAFGG 1169  
 Db 1087 -----PVEEYINQ-SVPRKPAQSVQNPYHNQPLHP-----APGRDLHYON--PHSN 1130  
 QY 1190 AVENPEYL-TPQGAAPQHPHPAPSPAFDNLVYWDQ-----DP-----PERGA 1232  
 Db 1131 AVSNPEYINTAQ-----PTLSSGFSBSALMIQKSHQMSLDNPDYQDFFPEAK 1181  
 QY 1233 PSTFTKGTPTAENPEYGLDVP 1254  
 Db 1182 PNGIFKG-PTAENPEYLRVAP 1202

RESULT 3  
 Q9EP98 ID Q9EP98 PRELIMINARY; PRT; 1210 AA.  
 AC Q9EP98;  
 DT 01-MAR-2001 (Tremblrel. 16, Created)  
 DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Epidermal growth factor receptor isoform 1.  
 GN EGFR.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C3H/101, 129/SVJ, AND 129/SVEYTAC;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Daniels A.J.,  
 RA Sinclair C.S., Pearse R.L., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,  
 RA Maile N.J.;  
 RA "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative Egr transcripts encoding truncated receptor  
 RT isoforms.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J;  
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Daniels A.J.,  
 RA Schehl C., Pearse R.L., Green P.J., Yee D., Lampland A.L.,  
 RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,  
 RA Maile N.J.;  
 RA "Comparative genomic sequence analysis and isolation of human and  
 RT mouse alternative Egr transcripts encoding truncated receptor  
 RT isoforms.";  
 RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF275366; AAC28045.1; JOINED.  
 DR EMBL; AF275365; AAC28045.1; JOINED.  
 DR EMBL; AF275367; AAC24386.1; -  
 DR HSSP; P1362; IFCK.  
 DR MGD; MGI:95294; Egfr.  
 DR InterPro; IPR000345; CytC\_heme\_bind.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR002290; Ser\_thr\_kinase.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR PRINTS; PR0109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SMO0261; Fu\_5.  
 DR SMART; SMO0220; S\_TKc; 1.  
 DR SMART; SMO0219; TyKc; 1.

DR PROSITE: PS00190; CYTOCHROME C; UNKNOWN 1.  
 DR PROSITE: PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE: PS00109; PROTEIN KINASE DCM; 1.  
 DR PROSITE: PS00109; PROTEIN KINASE TYR; 1.  
 KW ATP-binding; Receptor; Transferase.  
 SQ SEQUENCE 1210 AA; 134840 MW; 62CD021C9D32E18 CRC64;

Query Match 44.1%; Score 3004; DB 11; Length 1210;  
 Best Local Similarity 48.6%; Pred. No. 6 2e-221;  
 Matches 621; Conservative 166; Mismatches 368; Indels 120; Gaps 27;

QY 11 LLLALLPFGAA-STOYCTGTDMKRLPASPEHLMRLHLYOGCGVQGNLELYLPTN 68  
 DB 14 LLLALLPFGAA-STOYCTGTDMKRLPASPEHLMRLHLYOGCGVQGNLELYLPTN 73  
 QY 69 ASLSFLDIOGVGVYLAHNOYRQVPLQRLIVRGQULFEDNYALAVLDNGLPNTTP 128  
 DB 74 YDLSFLKIOGVGVYLAHNOYRQVPLQRLIVRGQULFEDNYALAVLDNGLPNTTP 124  
 QY 129 VTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYODTILMKDI---FKHNOALATLI 184  
 DB 125 -YGTNRGLRELPMRNLOEILIGAVRPSNNPILCNMDTIQMDIVQNVFMSNMSDL--- 180  
 QY 185 DTRSRACGECSPKCSRCWSSSEDCOSLTRVCAAGCA-RCKEPLPTDCHECCAAAG 243  
 DB 181 -QSHPSGCPKCDSPSCWGGGEGECCQLTKIICAOQCRRGRSRSDCCCHNCAAG 239  
 QY 244 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPEGRYTFGASCVTACPVNY 303  
 DB 240 CTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPEGRYTFGASCVTACPVNY 299  
 QY 304 LSTDVGSCTLVCPHNOEVAEDGTQRCCKSPKRCVAVCYGLMOYIKANSKFIETLE 363  
 DB 300 VVTDHSGCVACGCPDYEV-EEGDIRCKCKDGPCKVCKGICIGEFK-DTUSINATNK 357  
 QY 364 -FAGCKKIFGSLAFLEPSFDGDSANAPLOPELOVFELESLTYLYISAMPDLEPL 422  
 DB 358 HFRTCTAISGDLILPFAFGDSFTTRPPLDPELELTKVKTIGLILQAPDMDTUL 417  
 QY 423 SVFQNTQVIRGILHNGAVSLTQGLISWGLRSRELGSALILHNTHLCFVATVP 482  
 DB 418 HAFENLEIRGTRKQGFSLAVVGNITSLGRSLKEISDGVIIISGRNLCVANTIM 477  
 QY 483 DQFERNHOLLITARPDECEGEGLAGHQLCARHGWPGSTQCVNCSQIFRGECE 542  
 DB 478 KKLFGTPNQTKIMNNNAEKDCAVHNVHCNPLCSSGCGWPEPRDCVSCQNSRGECVE 537  
 QY 543 ECRVLQGLPREYVNAHCLPCHPECPONGSVTCFGEADQVACAHYKDPFCVARGPS 602  
 DB 538 KNILBEPREFEVNSECIGHECLPQAMNITCTGRPNQICQAHYIDGRHCVTICPA 597  
 QY 603 GVRPDLSTYMPIMKFPDEGACQCPINCHSCVDLDKGCAPQASPLTSLVSAVGL 662  
 DB 598 GIMGENNTL-VMKYADANNVCHLCHANCYGCAGPGLOGCEVWPSGPKPSIATGVGGI 656  
 QY 663 LVVVLGVVFGI-LIKRQCKIRKYTMRLLOETELVEPLTPSGAMPNOQMRILKETE 721  
 DB 657 LFTVV-VALGIGLFMRRIHVRKRLRLQERELVEPLTPSGEANOAHRLIKETE 715  
 QY 722 KAVYLGSGAFGNFTVSWLAVPKVSASHLEV---LRENTSPKANKETILDEAVYAGV 777  
 DB 716 KIRYLGSGAFG---TVYKGLMIPGEGVKIPVAIKELREATSPKANKETILDEAVYAGV 771  
 QY 778 GSPYVSRLLIGICTSTQVLTOLMPYGLLDHVENRGRIGSDILNMCQIAKGSYLE 837  
 DB 772 DNHRVCRLLIGICTSTQVLTOLMPYGLLDHVENRGRIGSDILNMCQIAKGSYLE 831  
 QY 838 DVSLVHRDLAARNVLYKSPNHVKTIFGLARLLDIDETEHADGKVP-KWMALESILRR 897  
 DB 832 DRLVHRDLAARNVLYKSPNHVKTIFGLARLLDIDETEHADGKVP-KWMALESILRR 891  
 QY 898 RFTHOSDVMSYGVTVWELMTFGAKPVDGIPAREIPLLELGEHLPPPICTIDVYIMTK 957

DB 892 IYTHOSDVMSYGVTVWELMTFGSKPYDGPASDISISILEKGERLPPPICTIDVYIMTK 951  
 QY 958 CMMLDSEKPRPRELVEFSMARDPQRFVYIQ-NEDLGPSPLDSTFYRLLEDDEDDGD 1016  
 DB 952 CMMLDSEKPRPRELVEFSMARDPQRFVYIQ-NEDLGPSPLDSTFYRLLEDDEDDGD 1011  
 QY 1017 LVDAEEYLVPOGCFPCDPAPAGAGVMHRRSSSTRSGGDLTLGLEPSEEEAPRSPLA 1076  
 DB 1012 VYDAEEYLVPOGCFPCDPAPAGAGVMHRRSSSTRSGGDLTLGLEPSEEEAPRSPLA 1037  
 QY 1077 PSEGAGSVFDDGLGMAKALQSLPHDPSPLORYSEDPVLPSET--DGYVAPLTC 1134  
 DB 1038 SLSATSN---NSVACINNGSCRVEDAFLLQRYSSDPPGAVTEDVIDDAFL----- 1087  
 QY 1135 PPEYVNOVDVAPPPPREPGLPAPAPAGATLERAKTLSPGKGVVVDVAFGAVENP 1194  
 DB 1088 PPEYVNOVDVAPPPPREPGLPAPAPAGATLERAKTLSPGKGVVVDVAFGAVENP 1136  
 QY 1195 EYL-TPGGAAPOPHPAPAFAPDNLVYWDQ-----DP-----BERGAPPSTF 1237  
 DB 1137 EYLNTAQ-----PCLSSGFNSPALMWIKGSHQMSLDNPDYQDPFKETKPNIGF 1187  
 QY 1238 KGTFAENPEYLGDLVP 1254  
 DB 1168 KG-PTAENAEYLRVAP 1203

## RESULT 4

QYTH40 PRELIMINARY; PRT; 1165 AA.  
 ID QYTH40  
 AC QYTH40;  
 DT 01-MAY-1999 (TREMBLrel. 10, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Receptor tyrosine kinase proto-oncogene.  
 GN XMRK.  
 OS Xiphophorus xiphidium.  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;  
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.  
 OX NCBI\_TaxID=8086;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RX MEDLINE=98241172; PubMed=9582016;  
 RA Dimitrijevic N., Winkler C., Welbrock C., Gomez A., Duschl J.,  
 RT "Activation of the Xmrk proto-oncogene of Xiphophorus by  
 RT overexpression and mutational alterations.";  
 RL Oncogene 16:1681-1690(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=RIO PURIFICATION;  
 RA Schartl M.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: U53471; AAD10500.2; -.  
 DR HSSP; P11362; IFGK.  
 DR InterPro; IPR000345; CytC\_heme\_bind.  
 DR InterPro; IPR000494; EGFR\_L\_domain.  
 DR InterPro; IPR000719; Euk\_kinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001899; Gram\_pos\_anchor.  
 DR InterPro; IPR001245; Tyr\_kinase.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Kinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 2.  
 DR PRINTS; PRO0109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_kinase; 1.  
 DR SMART; SM00261; Fu; 3.  
 DR SMART; SM00219; Tyk; 1.  
 DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE DCM; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.

DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00101; PROTEIN KINASE DOM; 2.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 KW ATP-binding; Kinase; Transferase; Tyrosine-protein kinase.  
 SQ SEQUENCE 1165 AA; 129614 MW; 7F7EE38D8771A74E CRC64;

Query Match 38.4%; Score 2617; DB 13; Length 1165;  
 Best Local Similarity 44.5%; Pred. No. 2, 6e-191;  
 Matches 571; Conservative 164; Mismatches 385; Indels 164; Gaps 35;

1 MELALCRWGLLALPPG-AAST---QVCTGDMKRLPASPEHRLDMLRYGQCV 55  
 4 LELL---LILLISIGCCSTDPKXVCGSTNQMM---LDNHYLKKMKMGSCV 56  
 QY 56 VQGNLEITVPTNASIFLQDIOEVQGYVLIANQVRQVPLQRLVIRGTQLEEDNYALA 115  
 DB 57 VLENLEITVQENQDLSFLQIOEVGYVLIANNEVSTIPLVNLRLRQNLVEGNFTLL 116  
 QY 116 VLNDGPIANTPTVGTASPGGLRELQRLTELKGGVLIQRMPCYQDITLMKDFRK 175  
 DB 117 VMSNYOK-NPSSP-DVYQVGLKQLQSLNLTSLSGVAKSHNPPLCNVETINMADIYDK 173  
 QY 176 NNOLATLIDITNSRACHPCSPCKGRWSESSDQSLTRTVCAQC-ARCKGFLPTD 234  
 DB 174 TSNPTMNLIPHAFERQCKDCPCVANGSCVAPGRHCQKPTKLCAEQCNRRCRGKPID 233  
 QY 235 CCEHQAAGCTGRKSHDCLALHFNHSGITELHCPALVTYNTDFESMRPGRITFGAS 294  
 DB 234 CCEHQAAGCTGRKSHDCLALHFNHSGITELHCPALVTYNTDFESMRPGRITFGAS 293  
 QY 295 CVTACPVNLTSTVSGCTTACPLHNOEVTAEQSTORCEKSKFCARVAGLGM---QYI 350  
 DB 294 CVCCECSNYVTE-GACVRCSCAGMLEVD-ENGRCKRCPDQVCPKVDIGIGISLNTI 351  
 QY 351 KANSKRTIGTELEFAGCKIFGSLAPLPSFDCDPASNTAPLOPELOVFTLEETGYL 410  
 DB 352 ANVSTNIG---SFSNCTKINGDIILNRNFEEDGPHYKIGPMPEHMLLTVEKITGYL 407  
 QY 411 YISAMPDLSLVSFNLQVIRGRIILHNGVS-LTIQGLISMLGRSLRELQSLALH 469  
 DB 408 VLMWPEPMSTLSVFCQLEITIRRTTSPRSFSTFVVQVSHQLWGLSSLEVSAGVILK 467  
 QY 470 HNTLFCVHTVMDQLFRNPHQALLHTANPEDECVGEGLAQHLCARGHCMGPGTQCV 529  
 DB 468 NTPQLRYASTINRRLFRSEDOQIEVDAR-----ENQTNNESSDEGCGPGTMCV 520  
 QY 530 NCSQFLRGQCEVCEKVLQGLPREYVNAHCLPCHPECCQNSVTCFGEADQCVACH 589  
 DB 521 SCHLVDRGGCVASCNLLQGEPRQAVDGRCVQCHOECIVQDTSITCYGGRPANCKAH 580  
 QY 590 YKDPFCVACRPGGVPRDLSYMPIMKFPDEGACOPCPINCTHSCVDLDDKGPAPQRAS 649  
 DB 581 FQDGPQCIIRCPHMLGSDTL-IMKADKMGCCQCHQNCCTGCGGPGISGRGP-IVS 638  
 QY 650 PLTSISAVVGLLVVLLGVFGLIKRROQKIRKTYMRRLLOETELVEPLTPSGAMPQ 709  
 DB 639 HSSLAVGLVSGLLITVALLIVLLRRRIK-RKRTIRRLLOEKELVEPLTPSGQAPQ 697  
 QY 710 AQRRLKETELKRVKVLGSGAFGNFTV-SFW-----LRPKXSASHLEKRENTSP 761  
 DB 698 AFRILKETEPKDRVLGSAFC---TYKGLMNDGGENIRLPAV---IKVLRATSP 749  
 QY 762 KAKEILDEAYNAGVSPYSRLGICLTSTVQVLTOLMPYGCCLLDHYENRGLSGSD 821  
 DB 750 KVAQVLDKAYVNASVDHPVCRLLGICLTSAVQLVTLQMPYGCCLLDHYENRGLSGSD 809  
 QY 822 LNMCMQIAKMGVLEEDVRLVHRLAARNTVKSPPHVKITDGLARLDDITETVHAQ 881  
 DB 810 LNMCMQIAKMGVLEEDVRLVHRLAARNTVKSPPHVKITDGLARLDDITETVHAQ 869  
 QY 882 GKVPKMALESILRRFTQSDVWSYGVTVLWELMTFGAKPYDIPAREIPDLLEGERL 941  
 DB 870 GKVPKMALESILRRFTQSDVWSYGVTVLWELMTFGAKPYDIPAREIPDLLEGERL 929

QY 942 POPPICTIDVYMIWKCMMIDSECRPREPRLVSEFSMARPPQRFVNIQNEDLGAPSLD 1001  
 DB 930 POPPICTIEVYMIILKCMIDPSSRPRLVSEFSMARPPQRFVNIQNEDLGAPSLD 986  
 QY 1002 STFYSLLEDDMGVLVNAERYLVQGFPCDPAPRAGGVNHHSSSTRSGGDLT 1061  
 DB 987 RRLFRLLSSDD-DVVADEYLL-----NYKRI-NRQGS----- 1018  
 QY 1062 GLEPSEEPAPRSPPLASEGASDVFDGLGMAKGLQSLPTHPSPQRYSEDPV-PL 1120  
 DB 1019 -----EPCIPRNGH-----FVRENSIALRYSIDPTQAL 1047  
 QY 1121 PSETDGYAPLTGSPQRYVNPQVPPQ-----PSPRE-----GPLP-AAPRAGATL 1167  
 DB 1048 EKDLDGH-----EYVQPGSETSSRLSDIYNPYEDLTDGMGPVSLSSQEAETNF 1097  
 QY 1168 ERAKTLSPGKVVDVAFGAVENPEYLTFCGGAARQPHPPAPFAPNLVYWDOP 1227  
 DB 1098 SRPEYLTNNQNSL---PLVSSGSDPDY---QAG-----YQAA-----L 1132  
 QY 1228 PERGAPSTFKGTPRAENPEYGL 1251  
 DB 1133 PQTGALTGNGMFLPAENLEYLGL 1156

RESULT 5  
 ID Q9MEF6 PRELIMINARY; PRT; 1137 AA.  
 AC Q9MEF6;  
 DT 01-NOV-1999 (TREMREL. 12. Created)  
 DT 01-NOV-1999 (TREMREL. 12. Last sequence update)  
 DT 01-JUN-2002 (TREMREL. 21. Last annotation update)  
 DE Receptor tyrosine kinase (fragment).  
 GN ERBB4.  
 OS Gallus gallus (Chicken).  
 OC Archosauria; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=HINDRAIN;  
 RX MEDLINE=99263203; PubMed=10328884;  
 RA Dixon M., Lumsden A., "Distribution of neuregulin-1 (nrg1) and erbB4 transcripts in embryonic chick hindbrain." Mol. Cell. Neurosci. 13:237-258 (1999).  
 RL EMBL; AF121963; AAD31764.1; -.  
 DR HSSP; P11362; 1FGK.  
 DR InterPro; IPR000494; EGFR\_L domain.  
 DR InterPro; IPR000719; Euk\_Pkinase.  
 DR InterPro; IPR002174; Furin-like.  
 DR InterPro; IPR001368; TNFR\_C6.  
 DR InterPro; IPR001245; Tyr\_Pkinase.  
 DR InterPro; IPR004019; YLP motif.  
 DR Pfam; PF00757; Furin-like; 1.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF01030; Recep\_L\_domain; 1.  
 DR Pfam; PF02757; YLP; 2.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Euk\_Pkinase; 1.  
 DR SMART; SM00261; FU; 3.  
 DR SMART; SM00219; Tyrc; 1.  
 DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE DOM; 1.  
 DR PROSITE; PS00109; PROTEIN KINASE TYR; 1.  
 DR PROSITE; PS00552; TNFR\_NGFR\_1; UNKNOWN\_1.  
 KW Kinase; Tyrosine-protein kinase.  
 FT NON TER 1  
 SQ SEQUENCE 1137 AA; 127927 MW; 4D616436F7DC4F CRC64;  
 Query Match 37.9%; Score 2580.5; DB 13; Length 1137;



Best Local Similarity 45.5%; Pred. No. 1.6e-188; Matches 521; Conservative 177; Mismatches 358; Indels 89; Gaps 29;

161 LCYODITLWKDIFHKNOALTLIDTRSPACHPCSMCKGSRGWSSSDCSLRTTV 220  
 Db 3 LCFADTTHMODIVNPMASNFLLVPTVSSGCGCHSCCTG-RGWSTENHCTLLKTV 61  
 QY 221 AGGC-ARCKGFLPDCCHGCACTGPKHSDCLACHFNHSGICEHCPALVTYNTDF 279  
 Db 62 AEQCDGRCYGPYVDCCHREACGSCSPKOTDCACMNFDSGACTVQCOTVNPPTF 121  
 QY 280 ESMNPEGRYTFGASCTYACPYNYLTDVGSCTLVCPHNOEVTADDTORCEKSPCA 339  
 Db 122 QLEHNHAKYTYGAFVCKKCPHNFV-VDSSCVACACSSKXEV-EENGIMCKRPTDICI 179  
 QY 340 RVYGLMGOVYKASKEFLGITELE-FAGCKKIFGSLFLDESPGCDASNTAPLOPQLO 398  
 Db 180 KACDGTGSL-VSAQTVDSNINIKFNCTKINGNLFLVTGHDGPHYHIAINPEKLN 238  
 QY 399 VFETLEETGYLYISAMPDLPDLVDFONTQVIRGRILHNGAVSLTQGLISWLGRL 458  
 Db 239 IFQVREITGYLNIQSPENMTDPRVFSNLVITIGRALYSGSLILKQGGITSLQFOSL 298  
 QY 459 RELSGSLALHNNHLCGVHTVHPMDQLFNPHQALHTARPEDECVGEGELACHOLCARG 518  
 Db 299 KQISAGNIYITDNNLCYHTVNTSLFSTPSQKTVHRRKKANCTADAMVCNELCSSD 358  
 QY 519 HCMGPGPTQCVCNCGQFLRGCEVECEVTLQGLPREYVNAHCLCPHECOP-ONGSVTCF 577  
 Db 359 GCKWGPDPQCLSCRFIRGRTICISCNLYGDFEFNANGSCMCDQCCKMDNMITY 418  
 QY 578 GPEADQVCAHAYDPPPCVARGSPGVKPLSYMPKFPDEBACOPCPINCTHSCVDL 637  
 Db 419 GPGDHCTKCFHFGDNCVCEKCDGLOGANSF--IFKYADEDECHPCPNCTOGGRGP 476  
 QY 638 DDGCG-----PAQRASPLTSIVSAV--GILLVAVLVGVFGILLKRRQKIRKY 685  
 Db 477 ASHCCTIYPTWRTOSTLQOHAR-TPL--IAAGVIGGLPIYIMGLTFVAYRRKIR-KKR 532  
 QY 686 TMRRLQETLEVEBLTPSGAMPNOQMRILKELTRKVKVLGSGAFENFTVSFWLRVP 745  
 Db 533 ALRRFL-ETELVEBLTPSGTAPNOAOLRILKELTEKVKVLGSGAF--TYVKGIWVP 587  
 QY 746 -----KVSASHLEVLRENTSPKAKELIDEAIVMAAGSPVSSLLGICLSTVQVLTOL 800  
 Db 588 EGETVKLPVA-IKILNETTGKAVNEFEDBALIASHMDHPLVLLDVCUSPTIQLTOL 646  
 QY 801 MPYGLLDHVENRGRGLGSDLLNMCQIAKMSYLEDVRLVHRDLAARVLYKSPNHVK 860  
 Db 647 MPOGLLDYVHEKDNIGSOLLNMCQIAKMYLDEBRRLVHRDLAARVLYKSPNHVK 706  
 QY 861 ITDFGLALDIDETEHADGKVPKIMMALESLRRRFTHQSDVMSVGTVMELMTFGA 920  
 Db 707 ITDFGLALDIDETEHADGKVPKIMMALESLRRRFTHQSDVMSVGTVMELMTFGG 766  
 QY 921 KPYGIPAREIPDLLEKGERLPORPICIYIMIMVCMIDSECRPRELVEFERSMA 980  
 Db 767 KPYGIPAREIPDLLEKGERLPORPICIYIMIMVCMIDSECRPRELVEFERSMA 826  
 QY 981 RDPORFVVIQNEP-LGPASPLDSTFYRSLLEDDMDGLVDAEYLVPQOQFCFDPAPGA 1039  
 Db 827 RDPORFVVIQNEP-LGPASPLDSTFYRSLLEDDMDGLVDAEYLVPQOQFCFDPAPGA 885  
 QY 1040 GGMVHHRSSSTSGGGDLTLGLEPSESEAPRS--PLAP-SEGAGSDVFGDGMGA 1096  
 Db 886 RTRIDSNRNGVYDGGYAAEGV-PMYRAPGCIIEAPVAQAQATEIFEDTCSGTLR 944  
 QY 1097 GLQSLFTHDPSPLQRYSEDPVPLPS-----ETDGVABLCGSPQEVYNOBPAPOP 1149  
 Db 945 KQVATLAKEDSSTQRYSDPTVFIPEYVIRGELDEDEGYMTMRKPTDYLNVPEENPV 1004  
 QY 1150 PSPREGPLPA-RPARGATLERAKTLSPKNGVVDV-----AGGAVENPEYLTPO 1200

Db 1005 SRKKNQDQAADVNPENH-----APNGQKAEDEYVNEPLVNTFANTLENAEYL--- 1054  
 QY 1201 GGAPOPHPPAPFSPAFNLYWDDPPERGA--PSTFKGPT-----AE 1244  
 Db 1055 -----KNLNPBKAQKAFNDPDYHNSLPPRSLQHPDYLOEISTYFKQNGRIRPIYAE 1109  
 QY 1245 NPEYL 1249  
 Db 1110 NPEYL 1114  
 RESULT 6  
 ID P79754 PRELIMINARY; PRT; 1328 AA.  
 AC P79754;  
 DT 01-MAY-1997 (TREMELREL. 03, Created)  
 DT 01-MAY-1997 (TREMELREL. 03, Last sequence update)  
 DT 01-JUN-2002 (TREMELREL. 21, Last annotation update)  
 DE EtdB3.  
 OS Fugu rubripes (Japanese pufferfish) (Takifugu rubripes).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;  
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;  
 OC Tetraodontidae; Takifugu.  
 NCBI\_TaxId=31033;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=9317347; PubMed=1007531;  
 RA Gellner K., Brenner S.  
 RT "Analysis of 148 kb of genomic DNA around the wnt1 locus of Fugu  
 rubripes".  
 RL Genome Res. 9:251-258(1999).  
 DR EMBL: AF056116; AAC34391.1; -.  
 DR HSSP: P11362; IFGR.  
 DR InterPro: IPR000494; EGFR L domain.  
 DR InterPro: IPR000719; Euk.pkinase.  
 DR InterPro: IPR002174; Furin-like.  
 DR InterPro: IPR001245; Tyr.pkinase.  
 DR Pfam: PF00757; Furin-like; 1.  
 DR Pfam: PF00069; pkinase; 1.  
 DR Pfam: PF01030; Recep.L.domain; 2.  
 DR ProDom: PD000001; Euk.pkinase; 1.  
 DR SMART: SM00261; FU; 3.  
 DR SMART: SM00219; TYRK; 1.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS00111; PROTEIN\_KINASE\_DOM; 1.  
 DR ATP-binding; Transferase.  
 SQ SEQUENCE 1328 AA; 148613 MW; A333039258B647B9 CRC64;  
 Query Match 32.3%; Score 2204; DB 13; Length 1328;  
 Best Local Similarity 39.2%; Pred. No. 1.4e-159; Matches 508; Conservative 161; Mismatches 418; Indels 208; Gaps 36;  
 9 WGLLALALPP--GAATQ-----VCTGDMKRLPASPETHLMDLRLHYOGCGVQGNIEL 62  
 Db 4 WRLLMCVASRLRAASSQGAQVCPQTQGLSTSGQENQVNLKDRYKCEIIMGNELEI 63  
 QY 63 TYPTNLSLSPDIOIEVQGYLIANQVRQVPLQRLVVRGTQFEDNYALAVDNGRP 122  
 Db 64 TQIESWDSFLKTRIEVGVYLIANNHFOELFLQQLRIRNSLYERFALSVPLN-- 120  
 QY 123 LNNVTVTASPGARELQRLSLTEILKGVLIQRNPOLCYODTILMKDIFHKNOALAT 182  
 Db 121 ----YKDG--FSLGNQGLMNLTELDSGVQIINNKYIRYGPWYWRIL--RNNDAPLE 173  
 QY 183 LIDTRNSRACHPCSPCKGSRGWSSSDCSLRTTVCAAGC-ARCKGFLPDCCHGCA 241  
 Db 174 ICFNGERGVCH--KSC-GNYCWGPGKQCCQILITVCAPOCNDRCFGTSPDCCHIEBA 229  
 QY 242 AGGTGKSHSDCLACHFNHSGICELHCPALVYNTDTPSMNPEGRYTFGASCTACPY 301  
 Db 230 AGCKGPLDTCACRLFNDSGACVPOCPTLLYKQTTOMETNPNAKYQYSGICVSGCPT 289

```

QY 302 NYLSTVSGCTLVCPHNDVETADGTQR-CEKSKPCARVCYGLMGOYIKANSKFIGIT 360
DB 290 HFV-VDSGSCVSVCPDMEV--ERGSORCELCISLCPKCEGTAE-----QRCVDS 342
QY 361 ELE-FAGCKKIFSLAFIPESFDGDPASNTAPLQPEOLOVEFEILEETITGVYISAPD 419
DB 343 NIDSFNCTKIQSLHFLVTGILGDPKVPVUDAKLEFRTVREITDILNQSFKEL 402
QY 420 PDLVFTQNLQVIRGILHNGAVSLTQGLISWLAGLSRELGSALIHNNTHLCFVT 479
DB 403 NDLVSFSLTTIGRSLFKRPSLMWRIFLTSLGLRSLREISDSGVYSQNAHLTYHT 462
QY 480 VPMDOLEFRNH-QALHTANRDECEVGEGLACHQICAGCHGCPPTQCVNCSQLEQ 538
DB 463 VMTQLEFRGSRVANSLSNRPAECVADGRVCDPLCSGCGPGPDCLCRNYSRG 522
QY 539 ECVEECRVLOGLPREYVNAH-CLPGHECOPONGSVTCFEPADOCVACAHKDPFCV 597
DB 523 TVAGCHFNNGIIPREFAGLVGVACHPECKPOTGASCTGPGADCMACTFRDPCY 582
QY 598 ARCPGKPDLSMPYWKPFDEGACQPCPTNCTHSCVDLDDKGCPRAPASPLISVA 657
DB 583 SSCPAGVN-DEGKGLIFKFNREHCEPCQCTGCGSGGLNDC--LEAARLTSSGQ 638
QY 658 VVGILLVVLGVVF-----GILKRRQOKIRKYMRLQETELVEPLTPSGAMPNA 710
DB 639 ITGIALGVPAGLIFCLVFLGLMHLRGLAIRKRAMRRLYESGSEFELPG-GEKITY 697
QY 711 QMELKTELKRVKVLGSGAFGNNFTVS--FWL-----RVKVASHLEVLRENTSPK 762
DB 698 HARLKPSDLRKIKPLGSGVFG---TVSKFWIPGEYTKIPVA---IKTIQSSGQ 749
QY 763 ANKEILDEAVMAGVSPYRSRLGICLTSTVOLVYQLMYGLLDHVENGRGLSQDL 822
DB 750 TTEETIDHLISKSLHPYVRLGLICPGCLQVLQVLSHSLHIIHQKSLIDPRL 809
QY 823 LNMCMQIAKMSYLEDVRLVHRDLAARNVLKSPNHYKITDFGLAELDIDETEHADG 882
DB 810 LNMCMQIAKMSYLEDVRLVHRDLAARNVLKNDYQVQISDYADVLVPRDKKYYSFT 869
QY 883 KVIKMALESILRRSFTHOSDVSQVYWEIMTGAQPYDGIIPAREIPDLIEKGERLP 942
DB 870 KPIKMALESILRRSFTHOSDVSQVYWEIMTGAQPYDGIIPAREIPDLIEKGERLP 929
QY 943 QPRICTIDYVMIWVKCMIDSECRPRELVSFSPMARPPQRFVYIQVEDLGPASPLS 1002
DB 930 QPRICTIDYVMIWVKCMIDSECRPRELVSFSPMARPPQRFVYIQVEDLGPASPLS 980
QY 1003 TFRSLEDDDDGDLVDAEYLVPOQGFPCPDAPAGGVHHRSSSTRSGGDLTG 1062
DB 981 -----EDSGMGEFL-----RGRSER---GILLEAD 1001
QY 1063 LEPSEEPARSPILAPBEGAGSDVFDGDLGMC--AAKGIQSLPTHPSLQ----- 1110
DB 1002 LEPSEEB-----GLGDRFAIPSLQPSFSWSTSSQINSYWMVQ 1040
QY 1111 -RYSDEPTVLPSEBTGYAFLPCSPQ- EYVNO-----PDVFPQPSPRE 1154
DB 1041 LRYD-----FAVSGGHIGYLPMSPSPYDTIRQLWYQSRSLSSVTLDRSAFRSSSE 1094
QY 1155 GFL--PAARPAKTLERAKTLPKNGVVKDVAFEGAVENPEYILTPGGAAPQPIPPA 1212
DB 1095 AELCEDGACGAGIFRVR-----FSSERGN-----FOG----- 1122
QY 1213 FSPAFNDLYWDCPPERGAPESTFKGTPTAENPE 1247
DB 1123 -----QQRKLTSTASSPSSFKTMADEDE 1146

```

```

DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Putative epidermal growth factor receptor (Fragment).
GN EGFR.
OS Anopheles gambiae (African malaria mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Eukaryota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae;
OC Anopheles.
OX NCBI_TaxID=7165;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SU4.
RA Lycett G.D.;
RT "Cloning, expression and localisation of the Anopheles gambiae
RT epidermal growth factor receptor.";
RL Submitted (NOV-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AJ301655; CAC35008.1;
DR HSRP; P13362; 1FGK.
DR InterPro; IPR000345; Cytochrome b5.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR000719; Euk_pkinase.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002290; Ser/thr_pkinase.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Kinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_pkinase; 1.
DR SMART; SM00261; Fu; 7.
DR SMART; SM00220; S_TKC; 1.
DR SMART; SM00219; Tyro; 1.
DR PROSITE; PS00190; CYTOCHROME C; UNKNOWN 4.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00117; PROTEIN KINASE DOM; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR Receptor.
FT NON TER.
SQ
SEQUENCE 1433 AA; 159585 MW; E3D9D88967724F07 CRG64;
Query Match 28.2%; Score 1923.5; DB 5; Length 1433;
Best Local Similarity 31.6%; Pred. No. 4.6e-138;
Matches 459; Conservative 197; Mismatches 400; Indels 397; Gaps 40;
QY 26 CTGDTMKRLPASPEPTLMDLRHLVQCGQVYQNLIELYPTNMSLSLFDIQEYQVYL 85
DB 1 CIGTNGKSVAPNREYHYKNLRDRTNCTVYDGNLEITWIONITDNLFIQHIRVYGYVL 60
QY 86 IAHNOVRQVPLORIRIVAGTQLF-----EDYVALAVLDNGPPLNNTTAVTGASPEGAREL 140
DB 61 ISLYDLPOVILPRLOITIGRTTFKLNKKEAYGLFV-----SFSHMTL 104
QY 141 QLRSLTELKSGVLIQRNPOLCYODTILMKDI-FHKNNQALTLIDTNRSACHGCSFMC 199
DB 105 ELPALRDLILGSGVGFENNYNLCMKMSIMBEILLAPQSMOYTNFSSPEVCPCHSC 164
QY 200 KGSRCWSSSSDQSLRTTVCAGCA--RCGSLPTDCHQCAAGCGPCHSCIALHL 257
DB 165 EVG-CWGGGANCCQFSLKNCSPQSGRCGPFRBCHFLCAGGCTGPPQSDCLACKN 223
QY 258 FNSGICELHCPALVYNTDTFESMPNEBGRYTGASCVTAQPNVYLSVDGSCGLVCP 317
DB 224 FYDDGVCKQCECPMQIYVPTNYFMEPNBDGAYAGATVRCF-EHLKMGACVCRKCP 282
QY 318 HNOETADGTQRCCKSKPCARVCYGLMGOYIRANSKFITELERFAGCKKIFGSLAFL 377
DB 283 GKMPQNS-----CVPCKGVCPKCPGEGI-----VHSDNTG-----NYKDCITIIIGSLIIL 329
QY 378 PESFDGDAASNT-----APLOEQLQVETLEETIEITGYIISAMPDPLSLVFOQLQ 429
DB 330 DQSFDDGQOVYVNTNFSFGRRYIKIDPDRLEIVSTYKEITGFINIQAHHPNFTLLNYFRNLE 389

```

```

QY 430 VIRGRLIHNGAY-SLTLOGIGISWILGRSLRSLGSLALIHNTHTLCHFVHTVPMQOLFEN 488
DB 390 VVGROJQKENTLFAVSIVYIKTSLKSLKSLKRVNSSVILENSDLCFEDDIDMSITKKS 449
QY 489 PHQALHTANPREDECEVGEGLACHOLCARGHCWGPPTOCVNCOSPLRGOCEVEBGRVLO 548
DB 450 SDHEVAVQKRNNTTECHBEGMECSBECOSKAGCKGKPEQCLEKKNVYKXKCLDSCK--- 506
QY 549 GLPREY-VNARHCLPCHRECPQONSVYTCGPEADOCVACAHKDPFCYACP----- 601
DB 507 SLPRLYSVDSKTCGDCHQCKD-----FCYGPNEDECGSMVYKDRFCVACBPTTKAM 561
QY 602 -----SGVKPDLSSYMBIKFPD----- 618
DB 562 NGTCINCKTCVCGRGRDTIADGCLSCDKALIGSDAKIERLMLMDSECPDQVSYDVL 621
QY 619 -EEG----- 621
DB 622 QEEGPKLQSLGKAVCRKCHRCCKCTGFGFHEQFCOECTGKKGEQCEDECPDYANEE 681
QY 622 -ACQPCINCT-----HSCVDL-----DP-----KCGPAEQ----- 646
DB 682 TRILPCHEQCRCHGLGDHHECRNLKLFESDPYDNATFTCVSNCPASHPKRPOEA 741
QY 647 -----RASPLTISVAVGILLVVLGVVFGI---LKKRQCKIRKRYTM 687
DB 742 GKIGPYCSADSMQSLRIEPOTVKIVGSMALILLCVFGALFVLFSHKKKKDAVVM 801
QY 688 RLLQETELVEPLTPSGAMPNOAMRLKETELKXVVGSGGFGPNNFVFSFVL----- 742
DB 802 TMLAGCESEPRPSVAVGNPLTLRIKEAERIRGVLWGAAG--RFFKGVWMEGES 859
QY 743 -RVPKVSASHLEVRENTSPKANKEILDEAVYVAGVSPVYSRLGICLTSTVQVLTOLM 801
DB 860 VKIPVA-----IKVLMEMSGSESSKEFLERAYIVASVHPMLKLKILAVCMSCMWLITQLM 915
QY 802 PYGCLDHRNRCRLSSQDLWMCQIAKMSLEBVLVLRVHDLAARNLVYSPNHVKI 861
DB 916 FLGCLDHYNNKOKIKSSKALLNWSTQIARGMALEBRLVHADLAARNLVYSPNCVKI 975
QY 862 TDFGLARLLDIDETEHADGKGVPIKMALESILRRFTHQSDVWSYGVTVWMLMTFGAK 921
DB 976 TVEGLAKLLDPSDEYRAGAKMFIKWLALBICIRHRYFTSKSDVMAGIITWELLITGAR 1035
QY 922 PYDGIPIAREIPDLLEKBERLPQPICTIDVYIMVWCWIDSECRPFRELVSEFSMAR 961
DB 1036 PYENVPAKDVPELIEIGHKLPQPDICSLDVYCIILSCWLDADARPTFKLAETFAEKAR 1095
QY 982 DPORFVYIQNEDLGPASPLDSTFYRSLLEDDMDGLV----- 1018
DB 1096 DPGRYLMI-----PGDKFMRLLPSTYNODEKDLIRTLAPVMAAAAAAAGASNVLD 1146
QY 1019 -----DAEELYLVQOGFFCPDPAAGAGVHNRHRSSTRSGGDLTLGLPSEEBAPR 1072
DB 1147 VPSTIAETDEYLOKTRPSIMLPGPSA-----VEPS-DEMPK 1182
QY 1073 S-----PLAP-----SEGAGSDVPFGDGLMGAKLQSLPTHDSPLQRYSEDPVLPSE 1123
DB 1183 SLRYCKDPLKPDDETDGHEKVV-----GVGSIR-----LNLPLD 1216
QY 1124 TDGYVAPLTCSPQPEYVNOPDVRPQPSPREGPLPAARPAATLERAKTSLPGKNGVVKD 1183
DB 1217 EDDYLMF-TQOSQ-----NQS-----TPG-----YMD 1237
QY 1184 VFARFGAVENPEYL-----TPGGAAQPHPPPAFSAFNNLYVMQDPPREKAPR 1234
DB 1238 LIGVAVADVNDPEYLMGSTQALAGLQSGM--PHTPP-----PP 1274
QY 1235 STFKGPTAENPE 1247
DB 1275 NTPNGMPTHQSHQ 1287

```

```

RESULT 8
ID Q9UK79 PRELIMINARY; PRT: 419 AA.
AC Q9UK79,
DT 01-MAR-2000 (TREMBLrel. 13, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Herstatin.
GN HER-2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99415951; PubMed=10485918;
RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;
RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted
RT autoinhibitor."
RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).
RN [2]
RP SEQUENCE FROM N.A.
RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF177761; AAD56009.2;
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR002174; Furin-Like.
DR Pfam; PF00757; Furin-Like; 1.
DR Pfam; PF01030; Recep_L_domain; 1.
DR SMART; SM00261; FU; 1.
SQ SEQUENCE 419 AA; 45472 MW; F6C1BE347E2D030C CRC64;

Query Match 27.5%; Score 1871; DB 4; Length 419;
Best Local Similarity 98.8%; Pred. No. 7.7e-135;
Matches 341; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MELALCRKGLLALLPFGAASVCTGTDMKRLPASPEHMDLRLHYOCQVVOGSL 60
DB 1 MELALCRKGLLALLPFGAASVCTGTDMKRLPASPEHMDLRLHYOCQVVOGSL 60
QY 61 ELTYLPTNASLSFLDIOEVQGVYIAHNOVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
DB 61 ELTYLPTNASLSFLDIOEVQGVYIAHNOVQVPLQRLIRVGTQLFEDNYALAVLDNG 120
QY 121 DPLANTTPYTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTIIMKDLFHKNOQLA 180
DB 121 DPLANTTPYTGASPGGLRELQRLSLTEILKGVLIQRNPOLCYQDTIIMKDLFHKNOQLA 180
QY 181 LTLIDNRRACHPCSPMKSGRCWGESESDCSLTRYACAGCARCKGPLPTDCHEQC 240
DB 181 LTLIDNRRACHPCSPMKSGRCWGESESDCSLTRYACAGCARCKGPLPTDCHEQC 240
QY 241 AAGCTGPKASDCLACLFHNSGICELACPAVTVNTDTFESMPNEGRYTFGASCVTACP 300
DB 241 AAGCTGPKASDCLACLFHNSGICELACPAVTVNTDTFESMPNEGRYTFGASCVTACP 300
QY 301 YNYLSDDVSCVLVQRLHNOETABDGTORCKSKRPARVYGL 345
DB 301 YNYLSDDVSCVLVQRLHNOETABDGTORCKSKRPARVYGL 345

RESULT 9
ID Q8R2X1 PRELIMINARY; PRT: 367 AA.
AC Q8R2X1,
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Hypothetical 40.2 kDa protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;

```

BACKFILE DOCUMENT INDEX SHEET

A DOCPHOENIX

APPL PARTS

IMIS Internal Misc. Paper  
LET Misc. Incoming Letter

371P PCT Papers in a 371 Application

A... Amendment including Elections

ABST Abstract

ADS Application Data Sheet

AF/D Affidavit or Exhibit Received

APPENDIX Appendix

ARTIFACT Artifact

BIB Bib Data Sheet

CLM Claim

COMPUTER Computer Program Listing

CRFL All CRF Papers for Backfile

DIST Terminal Disclaimer Filed

DRW Drawings

FOR Foreign Reference

FRPR Foreign Priority Papers

IDS IDS Including 1449

REV 7/02/03



Internal  
SRNT Examiner Search Notes  
CLMPTO PTO Prepared Complete Claim Set  
ECBOX Evidence Copy Box Identification  
WCLM Claim Worksheet  
WFEE Fee Worksheet

OUTGOING  
CTMS Misc. Office Action  
1449 Signed 1449  
892  
ABN Abandonment  
APDEC Board of Appeals Decision  
APEA Examiner Answer  
CTAV Count Advisory Action  
CTEQ Count Ex parte Quayle  
CTFR Count Final Rejection

NPL Non-Patent Literature  
OATH Oath or Declaration  
PET. Petition  
RETMAL Mail Returned by USPS  
SEQLIST Sequence Listing  
SPEC SPEC  
SPEC NO Specification Not in English  
TRNA Transmittal New Application

CTNF Count Non-Final  
CTRS Count Restriction  
EXIN Examiner Interview  
M903 DO/EO Acceptance  
M905 DO/EO Missing Requirement  
NFDR Formal Drawing Required  
NOA Notice of Allowance  
PETDEC Petition Decision

INCOMING

AP.B Appeal Brief  
C.AD Change of Address  
N/AP Notice of Appeal  
PA.. Change in Power of Attorney  
REM Applicant Remarks in Amendment  
XT/ Extension of Time filed separate

File Wrapper

FWCLM File Wrapper Claim  
IIFW File Wrapper Issue Information  
SRFW File Wrapper Search Info